

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 16:37:02 ; Search time 211.767 Seconds
(without alignments)

12680.960 Million cell updates/sec

Title: US-09-744-167-1

Perfect score: 4839
Sequence: 1 gcatactgaatcagcaggac.....atcaaaaaaaaaaaaaaaaaaa 4839

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	57.2	1.2	7218	1	US-08-232-463-14
C 2	46.4	1.0	1575	4	US-09-543-681A-615
C 3	45.8	0.9	429	4	US-09-543-681A-707
C 4	44.8	0.9	832	4	US-09-621-976-2813
C 5	42.6	0.9	30549	4	US-09-134-001C-322
C 6	42.4	0.9	202001	4	US-09-734-674-3
C 7	41	0.8	2515	4	US-08-221-017B-745
C 8	39.6	0.8	2981	1	US-08-257-073-2
C 9	39.6	0.8	2981	2	US-08-184-009-119
C 10	39.6	0.8	2981	2	US-08-458-356-119
C 11	39.6	0.8	2981	3	US-08-460-736-119
C 12	39.6	0.8	2981	4	US-09-535-370-119
C 13	39.6	0.8	3107	4	US-08-213-419B-1
C 14	39.6	0.8	6124	4	US-08-213-419B-3
C 15	39.6	0.8	786431	4	US-09-751-389-3
C 16	39.2	0.8	6297	3	US-09-045-201A-1
C 17	39.2	0.8	6297	4	US-09-619-062-1
C 18	38.6	0.8	3900	1	US-08-123-343A-6
C 19	38.6	0.8	10619	4	US-10-204-708-3
C 20	38.4	0.8	1882	1	US-08-257-073-12
C 21	38.4	0.8	1884	1	US-08-257-073-8
C 22	38.2	0.8	474	4	US-09-621-976-18033
C 23	38.2	0.8	2295	1	US-08-375-300-3
C 24	38.2	0.8	2295	3	US-09-177-431-3
C 25	38.2	0.8	2295	5	PCT-US95-16930-3
C 26	38.2	0.8	4080	1	US-08-375-300-1
C 27	38.2	0.8	4080	3	US-09-177-431-1

28	38.2	0.8	4080	5	PCT-US95-16930-1
29	38	0.8	832	4	US-09-621-976-2813
C 30	38	0.8	7786	4	US-09-790-988-2
C 31	37.6	0.8	918	3	US-09-105-390-35
C 32	37.6	0.8	1008	3	US-09-105-390-51
C 33	37.6	0.8	1949	3	US-09-105-390-2
C 34	37.6	0.8	2260	4	US-09-978-594-392
C 35	37.6	0.8	5136	4	US-08-956-171E-332
C 36	37.6	0.8	580073	4	US-08-545-528D-1
C 37	37.6	0.8	640681	4	US-09-790-988-1
C 38	37.6	0.8	1664976	4	US-08-916-421B-1
C 39	37.4	0.8	717	4	US-09-107-532A-2578
C 40	37.4	0.8	727	3	US-08-998-416-1011
C 41	37.2	0.8	400	4	US-08-956-171E-2481
C 42	37.2	0.8	3302	4	US-09-620-312D-475
C 43	37.2	0.8	5455	4	US-10-204-708-33
C 44	37	0.8	1871	4	US-08-956-171E-622
C 45	37	0.8	161652	4	US-09-497-855A-40

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOVLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 INMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZ9PT-Fis
US-08-232-463-14
Query Match 1.2%; Score 57.2; DB 1; Length 7218;

Best Local Similarity 10.5%; Pred. No. 5.6e-06;
Matches 50; Conservative 220; Mismatches 208; Indels 0; Gaps 0;

QY 436 ACAATGTCGATTGTAAGAAATGCTGTTGACAGAACGACGATTAATAAGAGAGAACTATAGT 495
DB 1446 AGAATTGCTACRR 1387
QY 496 TGGGATGATCAATGACGCTGTTGAACTGGGAGAGAAATGCGAAACCTGGCTTGT 555
DB 1386 RRR 1327
QY 556 CTGCAGATGAGAGAAATGTTCTGTTGACCGTCATGCAATACTGTGATAAAGACGA 615
DB 1326 RRR 1267
QY 616 TTCAAAACGATTTACAGGATGTAATAATTAATGATCAATCCCTTATGAGTCTTTT 675
DB 1266 RRR 1207
QY 676 AGCTGTTCACTGGATAATGAAAACAGACAACTGATCAATTTAGTTAGTATAAATGAG 735
DB 1206 RRR 1147
QY 736 TCCACTGAAAAGATATCAATTCAGAGAAACAAATGGATCCATTGAATAGACCCGAAACA 795
DB 1146 RRR 1087
QY 796 GAGGGAGATCTGTAACCATCTGTGCTTACTTCACTGATCTAGTCTAGCCAGTCTGT 855
DB 1086 RRR 1027
QY 856 TCCCTTCACAATTAAGAGTAGCGAGTAGTAGTAGAGACCCCTCCATGCTCGCA 913
DB 1026 TTCTGTGAGCGTAGGCAACGAGGAAATAGTTATAGTACCGCAGCTGATGGA 969

RESULT 2
US-09-543-681A-615
; Sequence 615, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 615
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-615

Query Match 1.0%; Score 46.4; DB 4; Length 1575;
Best Local Similarity 54.8%; Pred. No. 0.0034;
Matches 92; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 603 TGATAAAGGACATTACAAAACGATTTACAGGATTTGTAATAATTAATAGTCAATCCCT 662
DB 825 TGATAATGTTAACTTAGAAAATGATGTAATGTAATGTAATGTAATGTAATGTAATGTA 884
QY 663 TATGATGCTTTTACGCTGTTCACTGGATAATGAAAACAGACAACTGATCAATTTAGTTT 722
DB 885 AAATGATAATGTTAACTTAGAAAATGATAATTAATTAATTAATTAATTAATTAATTAAT 944
QY 723 TAGTATAATGATCCACTGAAAAGATATGTAATTCAGAGAACTATGAGAACTAT 770
DB 945 AAATATAAAGGACCTAATAAAGATATAAATAGAAATCTTATAT 992

RESULT 3
US-09-543-681A-707/c
; Sequence 707, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 707
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-707

Query Match 0.9%; Score 45.8; DB 4; Length 429;
Best Local Similarity 55.3%; Pred. No. 0.0019;
Matches 89; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 603 TGATAAAGGACATTACAAAACGATTTACAGGATTTGTAATAATTAATAGTCAATCCCT 662
DB 162 TCATAATGTTAACTTAGAAAATGATAATGTTAACTTAGAAAATGATAATGTTAACTTAGA 103
QY 663 TATGATGCTTTTACGCTGTTCACTGGATAATGAAAACAGACAACTGATCAATTTAGTTT 722
DB 102 AAATGATATGTTAACTTAGAAAATGATAATTAATTAATTAATTAATTAATTAATTAAT 43
QY 723 TAGTATAATGATGCTCACTGAAAAGATATGAATTCAGAGA 763
DB 42 AAATATAAAGGACCTAATAAAGATATAAATAGAAATA 2

RESULT 4
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match 0.9%; Score 44.8; DB 4; Length 832;
Best Local Similarity 10.1%; Pred. No. 0.0064;
Matches 34; Conservative 170; Mismatches 132; Indels 2; Gaps 1;

QY 3315 ATGGAAACTCCTCGGCTAAAGTATTCCTATCCGCTGATGTTGAGACTTGAGCTGA 3374
DB 360 TTTWKTTTWTYTTTTRMMWKKARRWYWKSTYACASRYKYTWGMYMKRMS- 302
QY 3375 ATATGACTTTATCCATGCCACTATTCAGTGTGATGTTTCGGAAGCCATTTTGGAGA 3434
DB 301 -TRWYCMWCKCMYRGRCAYTNARGMWSYANGKWSMRSMSCTMYKKGSTYW 243
QY 3435 GACGGGATACCATCATCATGTAATCTTCTTGGAGACTTCGAGAAATACCATGATACCTTCC 3494

Db 242 TMKCTCATWYWKYKRWKSTKWSGGRGGWNTSYSTRSYMTWASWMTMCMWGR 183
Qy 3495 AGTAGTTCAAGGTTTGGTGTGATGGAAGTTCGAAACCTAGCATCAAAATTCCTCAG 3554
Db 182 WWSYTWYMWGKKWRVYATTWRSAMWVWAAWMTMWYMWVWCMSSRGAAYRRRTMWVG 123
Qy 3555 CAACAGATACATGAGATGATGAAGCCATGAACAGTCCATGAGCATGTCTCTGSCAGG 3614
Db 122 YRYWRKKSYYRTFCWAYAWKTKRSYYWCWKKRWKRWKRWKRWKRWKRWKRWKRWK 63
Qy 3615 AGTGCTGCTTCAATGAAGAGGACAGACTCTCATCTTG 3652
Db 62 RYRWAWAWWVW 25

RESULT 5
US-09-134-001C-322
; Sequence 322, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 322
; LENGTH: 30549
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-322

Query Match 0.9%; Score 42.6; DB 4; Length 30549;
Best Local Similarity 47.9%; Pred. No. 0.47;
Matches 123; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Qy 1416 TGAGGACACTAATGGTATAGTGGAGGACAGTGTGGTGGATGGCAGATCGAGTCTAGA 1475
Db 26835 TGAATATGAACTAATCTATTGACAAATTCCTAGAAATCGTGAATCAAGCATTTGA 26894
Qy 1476 TTTAAAGGAACTTGCATTAGTGAAGTGAAGATGATTTCTCCACTGTATAGACAC 1535
Db 26895 TTTAAATGATCAATGAATCTTTAAGATGAATTAATCAAGCTTTTCTGTCCA 26954
Qy 1536 ACCAGCAGAAATATCTATATATGTTGTGATTCCTATGGAATGCAAGACCCAGTGT 1595
Db 26955 AGCAAGCTCAATATATATAATTCAGATGAAGATTTAAACACAAATTTGACCATGCTT 27014
Qy 1596 TTTCTTTTCCAAAGACTTTACCTCCAAAGAGATTCAGTACAGAGAAAGAAAT 1655
Db 27015 AAGTAATGCTCGAAAGTCTTTCGAAAGAAATGGTAAATAATTTAGTGAATAACAAAT 27074
Qy 1656 AGAGAAAGCAAGTCAG 1672
Db 27075 TCAGGACTCAACAG 27091

RESULT 6
US-09-734-674-3/c
; Sequence 3, Application US/09734674
; Patent No. 6498022
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001018

; CURRENT APPLICATION NUMBER: US/09/734,674
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 202001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(202001)
; OTHER INFORMATION: n = A,T,C or G
US-09-734-674-3

Query Match 0.9%; Score 42.4; DB 4; Length 202001;
Best Local Similarity 46.3%; Pred. No. 2.3;
Matches 139; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

Qy 532 AAGAAATGTGAAACCTGGTGTCTGCCAGATGAAGAAGATGTTCTTGTGTGACCCGTC 591
Db 21015 AAGCAATTTGGAACATATTTAAAGGAATCATTCAGAAAAATTTCCCTGATCTTTGGG 20956
Qy 592 ATGCATTAAGTGTGATAAAGGACATTACAAAACGATTTACAGGATTTAATAATTATAAT 651
Db 20955 AGTGAACATCCAGATATAAAGAAATTCAGAAAACACTTGCAGATACTATATAGATGAA 20896
Qy 652 AGTCAATCCCTTATGGATGCTTTTAGCTGTTTCACTGGATAATGAAAAACAGACAACTGAT 711
Db 20895 CATCATCAAGGCATATAGTCATCAGACTAACCAAGTCAATGCTAAAGAAAAATCTTAA 20836
Qy 712 CAATTTAGTTTGTAGTATAAAGTCCACTGAAAGATATCAATTCAGAGAAACAAATG 771
Db 20835 AGAGAGTTAGAGAGAACAGTCAAGTCACTGTTAAAGGAAATTCATCAGATGAGTGG 20776
Qy 772 GATCCATTGAATAGACCGAAAAACAGAGGGGAGATCTGTTAACCATCTGTGCTCTACTTCA 831
Db 20775 CTTTCTCAACAGAACCATATATAAGCCAAAAAGATAGGGGGCCTATTTTAGCCTTCTTGA 20716

RESULT 7
US-09-221-017B-745/c
; Sequence 745, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: ROSS, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 745:
SEQUENCE CHARACTERISTICS:
LENGTH: 2515 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc feature
LOCATION: 1...2515
US-09-221-017B-745

Query Match 0.8%; Score 41; DB 4; Length 2515;
Best Local Similarity 54.2%; Pred. No. 0.21;
Matches 83; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 398 CTATGGACAGACTGTAATCTAATCAGAGATGCCAGATGCGGATGATGATGATG 457
DB 2056 CTACCGGTCAGAGACTCTCTAATAAACAAGAGATGAAAAAACAAGATTAATGAATAC 1997
QY 458 CTGTTGCAAGACACAGTAAATTAAGAGAACTAGTTGGATGATCAATGCGAGTCTG 517
DB 1996 ATTATGAATTATCAGTGTCTTAAAGAGAACTTTTAAGGTAGTTGAATTTCAAGATA 1937
QY 518 TTGAAGTGGAGAGAGAAATGCGAACCTGG 550
DB 1936 TTGAAGTGGAGAGAGAAATGCGAACCTGG 1904

RESULT 8
US-08-257-073-2
Sequence 2, Application US/08257073
Patent No. 5766597
GENERAL INFORMATION:
APPLICANT: Paolletti, Enzo
APPLICANT: de Taisne, Charles
APPLICANT: Tine, John A.
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2981 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-2

Query Match 0.8%; Score 39.6; DB 1; Length 2981;
Best Local Similarity 49.1%; Pred. No. 0.65;
Matches 105; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 1427 ATGCTGATAGTGGAGACAGTGTGTTGGATTGGCAGATCGAGTCTAGATTAAAGGAA 1486
DB 1568 ATAAAGATAATAAAGGAAAGGTGCTGATACATGTTGATACAACTTAGAAAAAGAGATA 1627
QY 1487 CTTGCATTAGTCAAGTGAAGATGCTGTTCTCCACTGTTATAGACACACCCAGCAGCAA 1546
DB 1628 CTTATCATATGATCACTCAGATAATATGTTTGTAAATAAGAAATTTGTACACATTAA 1687
QY 1547 ATATCTATCTAATGGTGTGATGTTGATGGAATGCAAGACCCAGGTGTTCTTTTGTTC 1606
DB 1688 AAGATGAAATAATTTGATATCTAATCTCAAGTTGAAGATCAAGGTAATTTGTGATACTT 1747
QY 1607 CAAAGACATTACCTCCAAAGAGATTCAGTAAC 1640
DB 1748 CATGATTTTCTTCAAAATATCATTTAGAAAC 1781

RESULT 9
US-08-184-009-119
Sequence 119, Application US/08184009
Patent No. 5833975
GENERAL INFORMATION:
APPLICANT: Paolletti, Enzo
APPLICANT: Tartaglia, James
APPLICANT: Cox, William I.
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
NUMBER OF SEQUENCES: 217
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/184,009
FILING DATE: 19-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2530
TELECOMMUNICATION INFORMATION:


```
;
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066CURTMS
; INFORMATION FOR SEQ ID NO: 119:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2981 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-184-009-119

Query Match      0.8%; Score 39.6; DB 2; Length 2981;
Best Local Similarity 49.1%; Pred. No. 0.65;
Matches 105; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 1427 ATGGTGATAGTGGAGGACAGTGTGTTGGATTGGCAGATGCGAGGTCTAGATTAAAGGAA 1486
Db 1568 ATAAAGATAATAAGGAAAGGTGTCGTACATGTTGATACACACTTTAGAAAAGAGATA 1627
QY 1487 CTTGCATTAGTGAAGTGAAGAAATGTTGATTTCTCCACTGTTATAGACACACAGCAGCAA 1546
Db 1628 CTTTATCATATGATAACTCAGATAATATGTTTGTATAAAGAAATATTGTAACAGATTAA 1687
QY 1547 ATTATCTATCTAATGTTGATTCCTTATGGAATCGAAGCCAGCGTGTCTTTCTTTGTTTC 1606
Db 1688 AAGATGAAATAATATTGTATATCTAATCTTCAAGTTGAAGATCAAGGTAATTGTGATACTT 1747
QY 1607 CAAGACTTTTACCTCCAAAGAAAGATTCAAGTAAC 1640
Db 1748 CATGGATTTTCTTCAAAATATCATTTAGAAAC 1781

RESULT 10
US-08-458-356-119
; Sequence 119, Application US/08458356
; Patent No. 5942235
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458.356
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,009
; FILING DATE: 19-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066CURTMS
; INFORMATION FOR SEQ ID NO: 119:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2981 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-460-736-119

;
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066CURTMS
; INFORMATION FOR SEQ ID NO: 119:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2981 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-184-009-119

Query Match      0.8%; Score 39.6; DB 2; Length 2981;
Best Local Similarity 49.1%; Pred. No. 0.65;
Matches 105; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 1427 ATGGTGATAGTGGAGGACAGTGTGTTGGATTGGCAGATGCGAGGTCTAGATTAAAGGAA 1486
Db 1568 ATAAAGATAATAAGGAAAGGTGTCGTACATGTTGATACACACTTTAGAAAAGAGATA 1627
QY 1487 CTTGCATTAGTGAAGTGAAGAAATGTTGATTTCTCCACTGTTATAGACACACAGCAGCAA 1546
Db 1628 CTTTATCATATGATAACTCAGATAATATGTTTGTATAAAGAAATATTGTAACAGATTAA 1687
QY 1547 ATTATCTATCTAATGTTGATTCCTTATGGAATCGAAGCCAGCGTGTCTTTCTTTGTTTC 1606
Db 1688 AAGATGAAATAATATTGTATATCTAATCTTCAAGTTGAAGATCAAGGTAATTGTGATACTT 1747
QY 1607 CAAGACTTTTACCTCCAAAGAAAGATTCAAGTAAC 1640
Db 1748 CATGGATTTTCTTCAAAATATCATTTAGAAAC 1781

RESULT 11
US-08-460-736-119
; Sequence 119, Application US/08460736
; Patent No. 6265189
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,736
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,009
; FILING DATE: 19-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066CURTMS
; INFORMATION FOR SEQ ID NO: 119:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2981 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-460-736-119
```


; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6124
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
; US-08-213-419B-3

Search completed: May 18, 2004, 07:12:04
Job time : 217.767 secs

Query Match		0.8%	Score 39.6;	DB 4;	Length 6124;
Best Local Similarity		49.1%	Pred. No. 1.1;		
Matches	105;	Conservative	0;	Mismatches	109;
			Indels	0;	Gaps
					0;

QY	1427	ATGGTGATAGTGGAGGACAGTGTGTTGGATGGCAGATGCGAGTCTAGATTTAAAGGAA	1486
Db	4451	ATAAGATAATAAAGGAAAGGTGCTACATGTTGATACACTTTAGAAAAGAGATA	4510
QY	1487	CTTGCATTAGTGAAGTGAAGATGTGATTCTCCACTGTTATAGACACACAGCAGCAA	1546
Db	4511	CTTTATCATATGATACTCAGATAATATGTTTGTATATAAGATATTGTAACAGATTAA	4570
QY	1547	ATTATCTATCTAATGTTGTGATTCCTATGGAATGCAAGACCCAGGTGTTTCTTTGTTTC	1606
Db	4571	AAGATGAATAATAATTGATATCTAATCTTCAAGTTGAAGATCAAGGTAAATTGTGATACTT	4630
QY	1607	CAAGACTTTACCTCCAAAGAAGATTCAGTAAC	1640
Db	4631	CATGGATTTTGTCTCAAAATATCAITTAGAAAC	4664

RESULT 15
US-09-751-389-3
; Sequence 3, Application US/09751389
; Patent No. 6630334
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001067
; CURRENT APPLICATION NUMBER: US/09/751,389
; CURRENT FILING DATE: 2001-01-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 786431
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(786431)
; OTHER INFORMATION: n = A,T,C or G
US-09-751-389-3

Query Match		0.8%	Score 39.6;	DB 4;	Length 786431;
Best Local Similarity		52.4%	Pred. No. 45;		
Matches	87;	Conservative	0;	Mismatches	79;
			Indels	0;	Gaps
					0;

QY	1625	AAGAGATTCTAGTAACACAGAAAAGAAATAGAGGAAGCAAGTCAAGATGCTACTCAA	1684
Db	379503	AACGGGAAGAAGAAAAGAGAGAGAGAAAAGGAAAGAAAGAAAGAAAGAAAGAA	379562
QY	1685	ATATTTTGAACAGAGAGAAATGAGGCCACAGAGGAGTGGAAGTCTTTTAAACAGCA	1744

Db 379563 AGAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 379622
QY 1745 CTGGTGACCTTAATGAAGAAAATTTATTACATATTTTCTGTAGTCA 1790
Db 379623 AGAGGAAAGAAAGAAAGAAAGAAATGCCATATACAAATTTCTGTCTCTCA 379668

This Page Blank (uspto)

Result No.	Score	Query Match	Length	DB	ID	Description
1	1830	37.8	1830	9	US-09-764-878-355	Sequence 355, App
2	1830	37.8	1830	9	US-09-764-878-356	Sequence 356, App
3	1830	37.8	1830	15	US-10-079-854-355	Sequence 355, App
4	1830	37.8	1830	15	US-10-079-854-356	Sequence 356, App
5	620	12.8	703	15	US-10-066-543-681	Sequence 681, App
6	570.4	11.8	610	9	US-09-764-878-75	Sequence 75, Appl
7	570.4	11.8	610	15	US-10-079-854-75	Sequence 75, Appl
8	559	11.6	5402	15	US-10-079-977-11	Sequence 11, App
9	432	8.9	501	15	US-10-066-543-1215	Sequence 1215, App
10	401	8.3	401	15	US-10-066-543-1369	Sequence 1369, App
11	260.2	5.4	498	10	US-09-913-995-9986	Sequence 9986, App
12	220.8	4.6	969	16	US-10-264-237-394	Sequence 394, App
13	203.4	4.2	475	10	US-09-913-995-3724	Sequence 3724, App
14	154	3.2	437	9	US-09-983-965-4329	Sequence 4329, App

181	GGAAGAGGATCACTGTGCTTAATGACAGGACTGTAAATCTAAATCCAGAGATTGCCACAAT	240
441	GTGCAATTGATCAAAATCTGTTGCAGAGAGACCAGTTAAATTAAGAGAACTATAGTTGGGA	500
241	GTGCAATTGATGAAAATGCTGTTCAGAGAGACCAGTTAAATTAAGAGAACTATAGTTGGGA	300
501	TGATCAATGCAGTCTGTTGAAGTGGAGAGAGAAATGTGGAAACCTGCTGTCTGTGCC	560
301	TGATCAATGCAGTCTGTTGAAGTGGAGAGAGAAATGTGGAAACCTGCTGTCTGTGCC	360
561	AGATGAGAGAAATCTTCTGTTGTAGCCGTCATGCATAACTGTGATAAAAGGACATTACA	620
361	AGATGAGAGAAATCTTCTGTTGTAGCCGTCATGCATAACTGTGATAAAAGGACATTACA	420
621	AAACGATTTACAGATTGTAATAATTAATAATAGTCAATCCCTTAATGATGCTTTTAGCTG	680
421	AAACGATTTACAGATTGTAATAATTAATAATAGTCAATCCCTTAATGATGCTTTTAGCTG	480
681	TTCACTGGATAATGAAACACAGACAACTGATCAATTTAGTTTGTAGTATAAATGAGTCCAC	740
481	TTCACTGGATAATGAAACACAGACAACTGATCAATTTAGTTTGTAGTATAAATGAGTCCAC	540
741	TGAAAAGATATGAATTCAGAGAAACAAATGGATCCAATTGAATAGACCGAAAAACAGAGG	800
541	TGAAAAGATATGAATTCAGAGAAACAAATGGATCCAATTGAATAGACCGAAAAACAGAGG	600
801	GAGATCTGTTAAACCATCTGTGCTACTTCTCATCTGATAGTCTAGCCAGTGTCTGTCCCC	860
601	GAGATCTGTTAAACCATCTGTGCTACTTCTCATCTGATAGTCTAGCCAGTGTCTGTCCCC	660
861	TTCACAATTAAGGATGACCGGAAGTATAGGTAGAGACCCCTCCATGTCGCGAATTACAAG	920
661	TTCACAATTAAGGATGACCGGAAGTATAGGTAGAGACCCCTCCATGTCGCGAATTACAAG	720
921	TTTAAACGTTGATTCAGTAATCTCATCCACAGGAAACAGATGGAATGTCCTGCTTTAAAA	980
721	TTTAAACGTTGATTCAGTAATCTCATCCACAGGAAACAGATGGAATGTCCTGCTTTAAAA	780
981	GCAAGAGAACTATATACCAATGAGGACCTCACTGGCAAAATCAGCTCTCCTTAGGACAGA	1040
781	GCAAGAGAACTATATACCAATGAGGACCTCACTGGCAAAATCAGCTCTCCTTAGGACAGA	840
1041	TCTAGGAGTCCAAATTCCTTTTCCACATGAGTGAAGGGATTTGATGAAAAAGAGCC	1100
841	TCTAGGAGTCCAAATTCCTTTTCCACATGAGTGAAGGGATTTGATGAAAAAGAGCC	900
1101	AGCAGAGGAGACACACTGAAAGAACCCCTCCGGTCTGGTTTACCTTGTCTTCTCAAAAC	1160
901	AGCAGAGGAGACACACTGAAAGAACCCCTCCGGTCTGGTTTACCTTGTCTTCTCAAAAC	960
1161	AGACATGCCTTAATGGGTCCTGGAAGGAATAATGACTGTGAACCGGTGTTTCAGATTGCCTTGT	1220
961	AGACATGCCTTAATGGGTCCTGGAAGGAATAATGACTGTGAACCGGTGTTTCAGATTGCCTTGT	1020
1221	GCCTAATGAAGTTAGGGCTGATGAAGTAATGAAGGTTATGAACATGAAGAACTCTTGCCAC	1280
1021	GCCTAATGAAGTTAGGGCTGATGAAGTAATGAAGGTTATGAACATGAAGAACTCTTGCCAC	1080
1281	TACAGAAATTCCTTTAATATGACAGAGATTCTCTGAAATCTCAGGACATGACTAATTTGGAA	1340
1081	TACAGAAATTCCTTTAATATGACAGAGATTCTCTGAAATCTCAGGACATGACTAATTTGGAA	1140
1341	GTTGACTAACTAAATGAGATGAATGATAGCCAAAGTAAACGAAGAAAGAAAGTTTCT	1400
1141	GTTGACTAACTAAATGAGATGAATGATAGCCAAAGTAAACGAAGAAAGAAAGTTTCT	1200
1401	ACAGATTAGTCAGCCTGAGGACACTAATGTTGATAGTGGAGGACAGTGTCTTCGATTGGC	1460
1201	ACAGATTAGTCAGCCTGAGGACACTAATGTTGATAGTGGAGGACAGTGTCTTCGATTGGC	1260
1461	AGATGAGGCTTAGATTTTAAAGGAACTTGCATTAGTGAAGTGAAGATGTGATTTCTC	1520
1261	AGATGAGGCTTAGATTTTAAAGGAACTTGCATTAGTGAAGTGAAGATGTGATTTCTC	1320

```

RESULT 2
US-09-764-878-356 -
; Sequence 356, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 356
; LENGTH: 1830
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-356

```

	Query Match	37.8%;	Score 1830;	DB 9;	Length 1830;
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches 1830;	Conservative	0;	Mismatches	0;
				Indels	Gaps
201	AGATGAACAGCTTTCTTCTACTTTATTGGATACAAAGTGGGAATGAAGATTTCTAGATCTCC				260
1	AGATGAACAGCTTTCTTCTACTTTATTGGATACAAAGTGGGAATGAAGATTTCTAGATCTCC				60
261	TTCTCACCGCGTGCATTTAACCTACTTTGGCCAGTGTGAATCAATCTCTGCAGTTTCTAA				320
61	TTCTCACCGCGTGCATTTAACCTACTTTGGCCAGTGTGAATGAATCTCGAGTTTCTAA				120
321	TGAGTCACAACCAACATGAAGTCTTTCTCGCTGGCTCATTTCACTCCCTGACCAACAGA				380
121	TGAGTCACAACCAACATGAAGTCTTTCTCGCTGGCTCATTTCACTCCCTGACCAACAGA				180

381 GGAAGAGGATCACTGTGCTTAATGGACAGGACTGTATCTAAATCCAGAGATTGCCAAT 440
181 GGAAGAGGATCACTGTGCTTAATGGACAGGACTGTATCTAAATCCAGAGATTGCCAAT 240
441 GTGGATGATGAATGCTGTGTCAGAGACAGGATTAATTAAGAGAACTATAGTTGGGA 500
241 GTGGATGATGAATGCTGTGTCAGAGACAGGATTAATTAAGAGAACTATAGTTGGGA 300
501 TGATCAATGCACTGTGTTGAAGTGGGAGAGAGAAATGTGAAACCTGGCTGTCTGCC 560
301 TGATCAATGCACTGTGTTGAAGTGGGAGAGAGAAATGTGAAACCTGGCTGTCTGCC 360
561 AGATGAGAGATGCTGTGTTGTCAGCGTCATGATTAATCTGTGATAAAGACATTACA 620
361 AGATGAGAGATGCTGTGTTGTCAGCGTCATGATTAATCTGTGATAAAGACATTACA 420
621 AAACGATTTACAGGATTTGAATTAATTAATAGTCAATCCCTTATGGATGCTTTAGCTG 680
421 AAACGATTTACAGGATTTGAATTAATTAATAGTCAATCCCTTATGGATGCTTTAGCTG 480
681 TTCACTGGATTAATGAAGACAGACAACTGATCAATTTAGTTAGTATTAATGAGTCCAC 740
481 TTCACTGGATTAATGAAGACAGACAACTGATCAATTTAGTTAGTATTAATGAGTCCAC 540
741 TGAAGAGATGATCAATTCAGAGAAACAAATGGATCCATTAATAGACCGAAGACAGAGG 800
541 TGAAGAGATGATCAATTCAGAGAAACAAATGGATCCATTAATAGACCGAAGACAGAGG 600
801 GAGATCTGTTAACCATCTGTGCTCTACTTCATCTGATAGTCTAGCCAGTGTCTGTTCCCC 860
601 GAGATCTGTTAACCATCTGTGCTCTACTTCATCTGATAGTCTAGCCAGTGTCTGTTCCCC 660
861 TTCAAAATTAAGATGACCGAGATATAGTAGAGACCCCTCATCTGCGGATTAACAG 920
661 TTCAAAATTAAGATGACCGAGATATAGTAGAGACCCCTCATCTGCGGATTAACAG 720
921 TTTAACGGTTGATTCACTATCTCATCCAGGAAACAGATGGATGCTCTGTTAAAAA 980
721 TTTAACGGTTGATTCACTATCTCATCCAGGAAACAGATGGATGCTCTGTTAAAAA 780
981 GCAAGAGAACTATATACAGATGAGGACCTCATCTGGCAAAATCAGCTCTCTAGGACAGA 1040
781 GCAAGAGAACTATATACAGATGAGGACCTCATCTGGCAAAATCAGCTCTCTAGGACAGA 840
1041 TCTAGGAGTCCAAATCTCTTTTCCACATGAGTGGGATTTGATGAAAAAAGAGCC 1100
841 TCTAGGAGTCCAAATCTCTTTTCCACATGAGTGGGATTTGATGAAAAAAGAGCC 900
1101 AGCAGAGAGAGACCACTGAAGAACTCCCTCGGTCTGTTTACCTTTGCTTCAAAAC 1160
901 AGCAGAGAGAGACCACTGAAGAACTCCCTCGGTCTGTTTACCTTTGCTTCAAAAC 960
1161 AGACATGCTTAATGGTCTGGAAGGAATATGACTGTGACGGTGTTCAGATTGCCCTGT 1220
961 AGACATGCTTAATGGTCTGGAAGGAATATGACTGTGACGGTGTTCAGATTGCCCTGT 1020
1221 GCCTAATGAAGTGTAGGCTGATGAAATGAAGTTATGAACATGAAGAACTCTTGGCAC 1280
1021 GCCTAATGAAGTGTAGGCTGATGAAATGAAGTTATGAACATGAAGAACTCTTGGCAC 1080
1281 TACAGAACTCTTAATGACAGAGCAATTTCTGAACTCTCAGGACATGATTAATGGAA 1340
1081 TACAGAACTCTTAATGACAGAGCAATTTCTGAACTCTCAGGACATGATTAATGGAA 1140
1341 GTTGCACTAACTAATGATGATGATAGCAAGTAAACGAAGAAAGGAAAGTTTCT 1400
1141 GTTGCACTAACTAATGATGATGATAGCAAGTAAACGAAGAAAGGAAAGTTTCT 1200
1401 ACAGATTAGTCAGCTGAGGACACTAATGTTGATGATGAGGACAGTGTGTTGATGGC 1460
1201 ACAGATTAGTCAGCTGAGGACACTAATGTTGATGATGAGGACAGTGTGTTGATGGC 1260
1461 AGATGCAGGCTAGATTTAAAGGAACTTGCTGATTTAGTGAAGTGAAGATGTTCTTC 1520

1261 AGATGCAGGCTAGATTTAAAGGAACTTGCATTTAGTGAAGTGAAGATGTTCTTC 1320
1521 CACTGTTATAGACACACAGCAGCAAAATTTATCTATCTAATGTTGTTGATTCCTTAAGAA 1580
1321 CACTGTTATAGACACACAGCAGCAAAATTTATCTATCTAATGTTGTTGATTCCTTAAGAA 1380
1581 GCAAGACCCAGGTTGTTTCTTTGTTCCAAAGACTTTACCTCCCAAGAGATTCAGTAAC 1640
1381 GCAAGACCCAGGTTGTTTCTTTGTTCCAAAGACTTTACCTCCCAAGAGATTCAGTAAC 1440
1641 AGAAGAAAAAGAAATAGAGGAAAGCAAGTCAAGATGCTTCAAAATTTATTAAGACAGAG 1700
1441 AGAAGAAAAAGAAATAGAGGAAAGCAAGTCAAGATGCTTCAAAATTTATTAAGACAGAG 1500
1701 AGAAGAACGAGCCACAGAGGAGTGAAGTCTTTTAAACAGCACTGGTGAACCTTAAGAA 1760
1501 AGAAGAACGAGCCACAGAGGAGTGAAGTCTTTTAAACAGCACTGGTGAACCTTAAGAA 1560
1761 GAAAGAAATTTATACATAATTTCTGTAGTCAAGTTCCATCAGTGTGGGCAATCTTCCCC 1820
1561 GAAAGAAATTTATTAAGAAATTTCTGTAGTCAAGTTCCATCAGTGTGGGCAATCTTCCCC 1620
1821 CAAAGTACTAGCAAGCCTGCCATCTATCAGTGTCTTTTGGTGGTGAAGACCCCAAGCA 1880
1621 CAAAGTACTAGCAAGCCTGCCATCTATCAGTGTCTTTTGGTGGTGAAGACCCCAAGCA 1680
1881 ACCTTCTAATCTTAAGACTTCAAAATTTCCAAAGCCATTTACAGCACTTTACAAATGACTT 1940
1681 ACCTTCTAATCTTAAGACTTCAAAATTTCCAAAGCCATTTACAGCACTTTACAAATGACTT 1740
1941 TCCTCAAAACAGTGGAAATTAATCTTAAAAATTAATAATGATTTCTTGGGAAAGCAAAAT 2000
1741 TCCTCAAAACAGTGGAAATTAATCTTAAAAATTAATAATGATTTCTTGGGAAAGCAAAAT 1800
2001 AGGGGAAACTCAGCAACCAATGTATGCAG 2030
1801 AGGGGAAACTCAGCAACCAATGTATGCAG 1830

RESULT 3
US-10-079-854-355
; Sequence 355, Application US/10079854
; Publication No. US20030054368A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P4121C1
; CURRENT APPLICATION NUMBER: US/10/079,854
; CURRENT FILING DATE: 2002-02-22
; Prior Application removed - See File Wrapper for Palm
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 355
; LENGTH: 1830
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-079-854-355

Query Match 37.8%; Score 1830; DB 15; Length 1830;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1830; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

201 AGATGAAACAGTTTCTTCTTCTTATTTGATACAAAGTGAATTAAGATTCATAGATCCCC 260
1 AGATGAAACAGTTTCTTCTTCTTATTTGATACAAAGTGAATTAAGATTCATAGATCCCC 60
261 TTCTCAGCGCTGTCTATTTAAACCTTACTTTGCGCAGTGTGAATGAATCTGCAAGTTCTAA 320
61 TTCTCAGCGCTGTCTATTTAAACCTTACTTTGCGCAGTGTGAATGAATCTGCAAGTTCTAA 120
321 TGAGTCAACACCAACTGAAGTCTTCTCCTGGTCTATTGAGTCCCTGACCAAGCA 380

321 TGAGTCACACCAACTGAAAGTCTTCTCCCTGGCTCATTCAGCTCCCTCCCTGACACAGA 380
Db
121 TGAGTCACACCAACTGAAAGTCTTCTCCCTGGCTCATTCAGCTCCCTCCCTGACACAGA 180
Qy
381 GGAAGAGGATCACTGTGCTAAATGGAACAGACTGTAAATCTAAATCCAGAGATTGCCACAAAT 440
Db
181 GGAAGAGGATCACTGTGCTAAATGGAACAGACTGTAAATCTAAATCCAGAGATTGCCACAAAT 240
Qy
441 GTGGATTGATGAAATGCTGTGTCAGACAGACCAAGTTAAATTAAGAGAAAATATAGTTGGGA 500
Db
241 GTGGATTGATGAAATGCTGTGTCAGACAGACCAAGTTAAATTAAGAGAAAATATAGTTGGGA 300
Qy
501 TGATCAATGCACTGCTGTGTAAGTGGGAGAGAGAAATGTGAAACCTCGCTGTCTGCC 560
Db
301 TGATCAATGCACTGCTGTGTAAGTGGGAGAGAGAAATGTGAAACCTCGCTGTCTGCC 360
Qy
561 AGATGAGAGAAATGCTGTGTTAGTCGCTCATGCAATCTGTGTAAGAGAAATGCAATTA 620
Db
361 AGATGAGAGAAATGCTGTGTTAGTCGCTCATGCAATCTGTGTAAGAGAAATGCAATTA 420
Qy
621 AAACGATTTACAGGATTTGTAATTAATTAATAGTCAATCCCTTATGATGCTTTTAGCTG 680
Db
421 AAACGATTTACAGGATTTGTAATTAATTAATAGTCAATCCCTTATGATGCTTTTAGCTG 480
Qy
681 TTCACTGGATTAATGAAACAGACAACTGATCAATTTAGTTTATGATTAATAGTCCAC 740
Db
481 TTCACTGGATTAATGAAACAGACAACTGATCAATTTAGTTTATGATTAATAGTCCAC 540
Qy
741 TGAAGAAGATATCAATTCAGAGAAACAAATGGAATCCATTAAGACCCGAAACAGAGGG 800
Db
541 TGAAGAAGATATCAATTCAGAGAAACAAATGGAATCCATTAAGACCCGAAACAGAGGG 600
Qy
801 GAGATCTGTTAACCATCTGTCTCTACTTCAATCTGATAGTCTAGTCCAGTGTCTGTTCCCC 860
Db
601 GAGATCTGTTAACCATCTGTCTCTACTTCAATCTGATAGTCTAGTCCAGTGTCTGTTCCCC 660
Qy
861 TTCAATTAAGAGATCAAGGATAGTCTAGTCCAGTCCCTCATCTGCTGCGATTACAAG 920
Db
661 TTCAATTAAGAGATCAAGGATAGTCTAGTCCAGTCCCTCATCTGCTGCGATTACAAG 720
Qy
921 TTTAACGGTTGATTGAGTATCTCATCCAGGAAACAGATGATGCTCTGCTTTAAAAA 980
Db
721 TTTAACGGTTGATTGAGTATCTCATCCAGGAAACAGATGATGCTCTGCTTTAAAAA 780
Qy
981 GCAAGAGAACTATATACAGATAGGACCTCATCTGCAAAATCAGCTCTCTAGGACAGA 1040
Db
781 GCAAGAGAACTATATACAGATAGGACCTCATCTGCAAAATCAGCTCTCTAGGACAGA 840
Qy
1041 TCTAGGAGTCCAAATTCCTTTTCCCATGAGTGGGGATTTTGATGAAAAAGAGCC 1100
Db
841 TCTAGGAGTCCAAATTCCTTTTCCCATGAGTGGGGATTTTGATGAAAAAGAGCC 900
Qy
1101 AGCAGAGAGAGCACCCTCAAGAAATCCCTCCGGTCTGGTTTACCTTTGCTTCTCAACC 1160
Db
901 AGCAGAGAGAGCACCCTCAAGAAATCCCTCCGGTCTGGTTTACCTTTGCTTCTCAACC 960
Qy
1161 AGCATCCCTTAATGGTCTCGAAGGATATGATGTAAGGATGTAAGGATGTAAGGATGTAAG 1220
Db
961 AGCATCCCTTAATGGTCTCGAAGGATATGATGTAAGGATGTAAGGATGTAAGGATGTAAG 1020
Qy
1221 GCCTAATGAAGTTAGGCTGATGAAAAATGAAGTTATGAACATGAAGAAATCTTTGGCAC 1280
Db
1021 GCCTAATGAAGTTAGGCTGATGAAAAATGAAGTTATGAACATGAAGAAATCTTTGGCAC 1080
Qy
1281 TACAGAAATCTTAATGATGAGAGCAATTTCTCTGAAATCTCAGGACATGATTAATGGAA 1340
Db
1081 TACAGAAATCTTAATGATGAGAGCAATTTCTCTGAAATCTCAGGACATGATTAATGGAA 1140
Qy
1341 GTTCACATAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1400
Db
1141 GTTCACATAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
Qy
1401 ACAGATTAGTCAGCCTGAGGACACTAATGTTGATGATGAGGAGACAGTGTGTTGGATTGGC 1460

1201 ACAGATTAGTCAGCCTGAGGACACTAATGTTGATAGTGGAGGACAGTGTGTTGATTGGC 1260
Qy
1461 AGATGCAAGTCTAGATTATAAGGAACTTGCAATTAGTAAAGTGAAGAAATGTTGATTCTC 1520
Db
1261 AGATGCAAGTCTAGATTATAAGGAACTTGCAATTAGTAAAGTGAAGAAATGTTGATTCTC 1320
Qy
1521 CACTGTTATAGACACACAGCAGCAAAATTTCTATCTAATGTTGTTGATTCCTATGGAAT 1580
Db
1321 CACTGTTATAGACACACAGCAGCAAAATTTCTATCTAATGTTGTTGATTCCTATGGAAT 1380
Qy
1581 GCAAGACCCAGGTTCTTTTTCCTTCCAAAGACTTTACCTCCCAAGAGATTCAGTAAC 1640
Db
1381 GCAAGACCCAGGTTCTTTTTCCTTCCAAAGACTTTACCTCCCAAGAGATTCAGTAAC 1440
Qy
1641 AGAAGAAAAAGAAATAGAGGAAACAGCTCAGAAATGCTTACTCAAAATATTTATGAACAGAG 1700
Db
1441 AGAAGAAAAAGAAATAGAGGAAACAGCTCAGAAATGCTTACTCAAAATATTTATGAACAGAG 1500
Qy
1701 AGAAATGAGGCCACAGAGGAGTGGACTACTTTTAAACAGCACTGGTGACCTTAATGAA 1760
Db
1501 AGAAATGAGGCCACAGAGGAGTGGACTACTTTTAAACAGCACTGGTGACCTTAATGAA 1560
Qy
1761 GAAAAATTTATTTACATAATTTCTGTAGTCAAGTTCCATCAGTGTGGGCAATCTTCCCC 1820
Db
1561 GAAAAATTTATTTACATAATTTCTGTAGTCAAGTTCCATCAGTGTGGGCAATCTTCCCC 1620
Qy
1821 CAAGTGTAGCAAGCCCTGCCATCTATCAGTGTCTCTTTTGGTGGTGAAGACCCCAAGCA 1880
Db
1621 CAAGTGTAGCAAGCCCTGCCATCTATCAGTGTCTCTTTTGGTGGTGAAGACCCCAAGCA 1680
Qy
1881 ACCTTCTAATCTTAACTTCAAAATTCCAAGCCATTTATCAGACCATTTTACAAATGACTT 1940
Db
1681 ACCTTCTAATCTTAACTTCAAAATTCCAAGCCATTTATCAGACCATTTTACAAATGACTT 1740
Qy
1941 TCCTCAACACAGTGGAAATAATCTAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2000
Db
1741 TCCTCAACACAGTGGAAATAATCTAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1800
Qy
2001 AGGGAAAACTCAGCAACCAATGTATGCAG 2030
Db
1801 AGGGAAAACTCAGCAACCAATGTATGCAG 1830

RESULT 5

US-10-066-543-681
; Sequence 681, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Maxgarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 681
; LENGTH: 703
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

Db 61 AGTCTGTTCTTATCAGTGTGTGTAAGGGGAAAAAGGTTTTAAACAAAGTCTCTTAAAGTGGT 120
Qy 121 TTTTCCACCGAGGAGATTAATCTCCACGAGAGCTTACCACTGGGACAAAGGTGTTA 180
Db 121 TTTTCCACCGAGGAGATTAATCTCCACGAGAGCTTACCACTGGGACAAAGGTGTTA 180
Qy 181 GATGAATTTGAAACAAACGAGAGATGAACAGATTTCTTCTTACTTATTTAGTGAACAAAGTGG 240
Db 181 GATGAATTTGAAACAAACGAGAGATGAACAGATTTCTTCTTACTTATTTAGTGAACAAAGTGG 240
Qy 241 AATAAGATTCAGATCCCCCTTCTCACCGCTGTCAATTAACCTACTTTGGCCAGTGG 300
Db 241 AATAAGATTCAGATCCCCCTTCTCACCGCTGTCAATTAACCTACTTTGGCCAGTGG 300
Qy 301 AATGAATCTCAGTTCTTAATGAGTCAACACCAACTGAAAGTCTTCTCCCTGGCTCAT 360
Db 301 AATGAATCTCAGTTCTTAATGAGTCAACACCAACTGAAAGTCTTCTCCCTGGCTCAT 360
Qy 361 TCAGTCCCTGACCCACAGAGAGAGATCACTGTGCTTAATGGACAGACTGTATCTA 420
Db 361 TCAGTCCCTGACCCACAGAGAGAGATCACTGTGCTTAATGGACAGACTGTATCTA 420
Qy 421 AATCCAGAGATTCGCAATGTGGATTGATGAATAATGCTGTTGCAGAGACCACTTAAT 480
Db 421 AATCCAGAGATTCGCAATGTGGATTGATGAATAATGCTGTTGCAGAGACCACTTAAT 480
Qy 481 AAGAGAACTATAGTTGGGATGATCAATGAGTGTGTTGAAGTGGGAGAGAAATGT 540
Db 481 AAGAGAACTATAGTTGGGATGATCAATGAGTGTGTTGAAGTGGGAGAGAAATGT 540
Qy 541 GGAACCTGCTGTGTCGAGATGAGAGAAATGCTTCTGTCAGCCGTCATGCATA 598
Db 541 GG-AACTGCTGTGTCGAGATGAGAGAAAG-TCCTGTGTAGCCGTCATGCATA 596

RESULT 8

US-10-072-977-11
; Sequence 11, Application US/10072977
; Publication No. US20030152933A1
; GENERAL INFORMATION:
; APPLICANT: Barash et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PT057P1
; CURRENT APPLICATION NUMBER: US/10/072,977
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: PCT/US01/25288
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: US 60/225,215
; PRIOR FILING DATE: 2000-08-14
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 11
; LENGTH: 5402
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-072-977-11

Query Match 11.6%; Score 559; DB 15; Length 5402;
Best Local Similarity 60.8%; Pred. No. 5e-142;
Matches 989; Conservative 0; Mismatches 620; Indels 18; Gaps 4;
Qy 2824 CTTATTCCTGAAGATGGCTTCTCCCACTTCTCATCTCCAGTGTGTAAAGG---AGAC 2880
Db 3157 CCTAATGATGAGGACAGTTTCCCTCCACTTCTGGTGTGATCTGGAGAAAGGGATCAGTG 3216
Qy 2891 TATGCTGGAGAGAGAACCATCATCAGATTTTCAGTAATGACAGATTTGGAGGATGGTGGC 2940
Db 3217 CCTGTAGTAGAAGAACATCCATCTCATGAGCAGATCATTTTCTTCTTGAAGGTGAAGC 3276
Qy 2941 CCTGACCCACATGTATTTGTTTAAATGCAATTTGTTGCAATGGTTTAAATTTGTAAT 3000
Db 3277 TTTATCTCTGTACATTTGTCCTAAATGCTTAATCTACTCGTGAATGTCATATTTT 3336

Qy 3001 TATGTGAACAGGAAGTGTGTGTGTTTCAACACCAAGGAATGCATGCAGTGGTCACTCT 3060
Db 3337 TATTCCTCAGACAAATATTTGGTACTTTTCAACCAATGATTTGCATGGCTTGGGACAGCA 3396
Qy 3061 GAGATAGTCACTTCTTACAGTGTTTTACCGGATGAAAGTGTTCGCCAAAGGATATCTTT 3120
Db 3397 GAAATATATTTCTTATTTGTTATTTTCCAAATGAGATACTATTCCTTAAGGACATCTTC 3456
Qy 3121 AATCACTTTTGTGCAGCTTTATCGGGATGCTCTGGCAGGGAATGTGGTGAACACTTGGGA 3180
Db 3457 AGACTATTTTATCACCATATATAAGGATGCTCTAAAGGAAATACATAGAAATCTTGAC 3516
Qy 3181 CATTCCTTCTTCAAGTCAAGTTCCTTGGCAGTAAAGAACATGTGTGATTTCTTATATGTG 3240
Db 3517 AATATTAATTTTACTGAGATTTTCTCAGTAGCAAGGATCAGGAGGATTTCTGTTTAT 3576
Qy 3241 ACATCTACTACTACAGTCACTGCAAGACCTAGTACTCCCAACCCCACTTACTTGTGGG 3300
Db 3577 ACACCTACTTTTCAGAAACTTTGATGATCTCTCAITACCAAGTAATCTTTCTTTGTGGA 3636
Qy 3301 ATTCTTATCCAGAAATGGGAACTCTTGGGCTAAAGTATTTCTTATCCGCTGATGTTG 3360
Db 3637 ATTCTTATCCAGAACTTTGAGATTCCTCGGGCAAGGTTTTTCTTATGCGTTTAAATG 3696
Qy 3361 AGACTTGAGTGAATATCGACTTTTATCCATGCCCACTTATTCAGTGTGATTTTGGGAAG 3420
Db 3697 AGATTGGGTGCAGATATAAGCCATATCTGCTCTCTAACCAAGCATCAGAGGCCGAAA 3756
Qy 3421 CCATTTGTTGGAGAGACGGGGATACCATCATGATCTCTTTCAGAGCTTCAGAAATTAC 3480
Db 3757 CCTCTTTTGGAGAAATAGGACACACTATTAAGAACTTACTTGTGACCTTCGAAATTAC 3816
Qy 3481 CAGTATACCTTCCAGTAGTTTCAAGTTTGGTGGTGTGATATGGAAGTTTCGGAATACTAGC 3540
Db 3817 CAGTATACCTTGCATATAATAGATCAACTGTTGATTCATATGGAATGGGAAAGTGC 3876
Qy 3541 ATCAAAATTTCCAGCAACAGATACAAATGAGATGATGAAGCCATGAACAGTCCATGAG 3600
Db 3877 ATAAAAATACCACCGGAAAAAGTACAGTGTATTAATGAAGTACTAAATTTCTTCCAATGAG 3936
Qy 3601 CATGTCCTGGCAGAGGTGCCTCTTCAATGAAGAGGAGAGCTCTCATCTCTGTGTGTA 3660
Db 3937 CATGTCATTTAGCATTTGGAGCAAGTTTTCAGTACAGAGCAGATTTCTCATCTAGTGTATA 3996
Qy 3661 CAGAATGATGATGGAACCTATCAGACCCAGGCTATCAGTATTTCACAATCAGCCCCAGAAA 3720
Db 3997 CAGA---ATGATGGAATTTATGAACACAGGCCAACAGTCCCACTGGCCATCTAGAAAA 4053
Qy 3721 GTGACTGTGCGCAGTTTCTTTGTGTTCAGTGGCGCTCTGAAATCTCTCTCTGATACCTT 3780
Db 4054 GTGACAGGTGCAAGTTTGTGTTTCAATGAGGCTCTAAAAAACATCTTTCAGGATTTCTT 4113
Qy 3781 GCCAAGTCCAGTATTTGTGAAGATGTTTATGTTCCAGATTACTGCAGAGAACATGGAT 3840
Db 4114 GCTAAGTCCAGCATAGTTGAAGTGGCTTAATGGTACAAATAACTCCAGAGACCATGAAT 4173
Qy 3841 TCCTTGAGCAGGCACTGCGAGAGATGAAGGACTTCACCATCCCTGTGGGAGCGGAC 3900
Db 4174 GGCTTCCGCTAGCTTTACGAGAACAGAAAGACTTTTAAATTTACATGTGGGAAAGTTGAT 4233
Qy 3901 GCGGAGAACCCAGGAGCAGATCCACATCCAGTGGTGGATGATGACAAAGAACGTTAGC 3960
Db 4234 GCAGTAGACCTGAGAGAAATACGTGGATATCTGCTGGGTAGATGCTGAGAGAAAGGAAAC 4293
Qy 3961 AAGGGTGTCTAGTCTTATAGTGGAAAGTCCATGGAGACTATAACAAATGTGAAGATA 4020
Db 4294 AAGAGGATTTACGTTTCAGTGGATGATATCATTTACAAGGATTTTCCAAAGTGAATAATA 4353
Qy 4021 TTCATGGATCAGAAATATAAGCAATGGAAGTAATCAGATGGAACAGAGGTGTTTTTC 4080
Db 4354 AAATCGAAGCAGATTTTGAACCCGATGAGAGATTTGTAATATGTTACCGAGGTGTTCTAC 4413
Qy 4081 CTAGAAAAAGATGACCCAGCAATTTGCCTCAGTGTATCTCTGAGATCAGATGATGACT 4140

QY 609 AAGGACATTACAAACGATTTACAGGATTTCTAATATTAATATTAATAGTCAATCCCTTATGGA 668
DB 161 AAGGACATTACAAACGATTTACAGGATTTCTAATATTAATATTAATAGTCAATCCCTTATGGA 102
QY 669 TGCTTTTAGCTGTTCACTGATATGAAGAAACAGACAACTGATCAATTTAGTTTATGAT 728
DB 101 TGCTTTTAGCTGTTCACTGATATGAAGAAACAGACAACTGATCAATTTAGTTTATGAT 42
QY 729 AATGAGTCCACTGAAAGAAATGATGAATTCAGAGAAACAAA 769
DB 41 AATGAGTCCACTGAAAGAAATGATGAATTCAGAGAAACAAA 1

RESULT 11
US-09-918-995-9986
; Sequence 9986, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9986
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(498)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-9986

Query Match 5.4%; Score 260.2; DB 10; Length 498;
Best Local Similarity 97.1%; Pred. No. 1.8e-60;
Matches 265; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 287 CTTTGGCCAGTGGAATGATCTCGAGTTCTTAATAGTCACACCACTGAAAGTCT 346
DB 226 CTTGCGCCAGTGGAATGATCTCGAGTTCTTAATAGTCACACCACTGAAAGTCT 285
QY 347 TCTCCCTGGCTCATTGAGTCCCTGACAGAGGAGGATCACTGTGCTAATGGAC 406
DB 286 TCTCCCTGGCTCATTGAGTCCCTGACAGAGGAGGATCACTGTGCTAATGGAC 345
QY 407 AGGACTGTAATCTAAATCCAGAGATTGCCAAATGTGGATTGAAATGCTGTGCGAG 466
DB 346 AGGACTGTAATCTAAATCCAGAGATTGCCAAATGTGGATTGAAATGCTGTGCGAG 405
QY 467 AAGACCAGTTAATTAAGGAACTATAGTGGGATGATCAATGAGTGTGGAAGTGG 526
DB 406 AAGACCAGTTAATTAAGGAACTATAGTGGGATGATCAATGAGTGTGGAAGTGG 465
QY 527 GAGAGAAAGAAATGGAACCTGGCTGTCTGC 559
DB 466 GAGAGAAAGAAATGGAACCTGGCTGTCTGC 498

RESULT 12
US-10-264-237-394
; Sequence 394, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131P1
; CURRENT APPLICATION NUMBER: US/10/264,237

; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 394
; LENGTH: 969
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (845)...(845)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (871)...(871)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (897)...(897)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (899)...(899)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-237-394

Query Match 4.6%; Score 220.8; DB 16; Length 969;
Best Local Similarity 59.4%; Pred. No. 2.1e-49;
Matches 442; Conservative 1; Mismatches 292; Indels 9; Gaps 4;

QY 2824 CTTATTCTTGAAGATGGCTTCCTCCCATTCATCTCCACGTGGTGTAAAGG---AGAC 2880
DB 202 CCTAATGATGAGGACAGATTTGGCCCACTCTCTGGTTGCATCTGGAGAAAGGATCAGTG 261
QY 2881 TATGCTGTGGAAGAAACCATCAGATTTTCAGTAATGCACAGTTGGAGGATGTTGTC 2940
DB 262 CTGTAGTAGAAGAACATCCATCTCATGAGCAGATCAITTTCTCTTGAAGTGAAGC 321
QY 2941 CTGACCCACTTGTATTTGTTTAAATGCAAAATTTGTGTCATGTTTAAATGTTAAAT 3000
DB 322 TTTTCATCTCTGTTACATTTGTCCTAAATGCTAATCTACTCGTGAATGTCAATTCATATT 381
QY 3001 TATGTGAACAGAAAGTGTGGTGTTCACCAAGGAGGATGATGCAGTGGTGGTCACTCT 3060
DB 382 TAT-TCCTCAGACAAATATGGTACTTTTCAACCAATGGATTGCGGCTGGGACAGG--- 437
QY 3061 GAGATAGTCAATCTCTACAGTGTTCACCGGATGAAAGTGTTCGCAAGGATATCTTT 3120
DB 438 CAGAAATATATCTATTGTTATGTTTGCCTAATGAGATACTATTCTTAAGGACATCTTC 497
QY 3121 AATCACTTTGTGAGCTTTTATCGGATGCTCTGGCAGGGAATGTGGTAGCAACTTGGGA 3180
DB 498 AGACTATTTTATCACCATAATAAGGATGCTCTTAAAGGAAAAATACATAGAAAACCTTGGAC 557
QY 3181 CATTCCTCTTTCAGTCAAAAGTTTCTTGGCAGTAAGAACATGTTGATCTTATATATGTG 3240
DB 558 AATATTACCTTTACTGAGAGTTTCTCAGTAGCAGGATCACGAGGATTCCTGTTATT 617
QY 3241 ACATCTACCTTACCAGTCACTGCAAGACCTAGTACTCTCCAAACCCCACTTACTGTTTGGG 3300
DB 618 ACACCTACTTTTCAGAAACTTGATGATCTCTCATTTACCAAGTAATCTTTCTTTGTGGA 677
QY 3301 ATTCTTATCCAGAAATGGGAAACTCTTGGGCTTAAGATATTTTCTTCCATCCGCTGATGTTG 3360
DB 678 ATTCTTATCCAGAACTTGAGATTTCCCTGGGCAAGGTTTTTCTTATGCGTTAATGTTG 737
QY 3361 AGACTTGGAGCTGAATATGACCTTTTATCCATGCCACTATTCAGTGTTCAGATTCGGAAG 3420
DB 738 AGATTGGTGCAGAAATATAAGCATATCTGCTCTCTAAGAGCATCAGAGGCGGAAAA 797

QY 3421 CCATTGTTGGAGACGGGCATACCATCATGTAATCTCTTGACAGCTTCAGAAATTAC 3480
Db |||||
798 CTTCTTTTGGAGAAATAGGACACACTATTATGAACCTTACCTTGTGANCCTCGAAATTAC 857
QY 3481 CAGTATACCTGCCAGTAGTTCAGGTTTGGTG--TTGATATGAAGTTCCGGAACCTA 3538
Db |||||
858 CAGTATACCTGCTNTTATATAAATCAACTGTGAATCCNTTGGAAATGGGAAAAAACC 917
QY 3539 GCATCAAAATCCCAAGCACAGAT 3562
Db 918 GCCTTAAATCCCGGAAAGTT 941

RESULT 13
US-09-918-995-3724
; Sequence 3724, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3724
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-3724

Query Match 4.2%; Score 203.4; DB 10; Length 475;
Best Local Similarity 64.6%; Pred. No. 7.9e-45;
Matches 303; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 3025 TTCACACCAAGGAATGATCGAGTGGTCTGAGTAGTCATCTTCTACAGTGT 3084
Db |||||
3 TTTTCAACCAATGGATTCATCGTGGTGGACAGCAGCAAAATATATCTTATGTTATGT 62
QY 3085 TTACGGATGAAAAGTGTTCCTCCAAAGGATATCTTTAATCATCTTGTGCAGCTTTATCGG 3144
Db |||||
63 TTGCCAATGAAGTACTATTCTTAAGGACATCTTCAGACTATTATCACCATATATAAG 122
QY 3145 GATGCTTGGCAGGAATGCTGGAGCACTTGGGACATTCCTTCTTCAGTCAAAGTTTC 3204
Db |||||
123 GATGCTCTAAAGGAAAAATACATAGAAAACCTTGGACAAATATTACCTTTTACTGAGAGTTT 182
QY 3205 CTTCGCTAAGAACATGTTGGATCTTATATGTGACATCTACCTACAGTCACTGCAA 3264
Db |||||
183 CTGATGACAGGATACAGGAGGATCTGTTTATACACTACTTTTCAGAACTTGAT 242
QY 3265 GACCTAGTACTCCAAACCCACCTTATCTTGTGGGATCTTATCCAGAAATCGGAAACT 3324
Db |||||
243 GATCTCTCATPACCAAGTAATCTCTTTTGTGGAATCTTATCCAGAAAGCTTGAGATT 302
QY 3325 CTTTGGGCTAAAGTATTTCTCTATCCGCTGATGTTGAGCTTGGAGCTGAATTCGACTT 3384
Db |||||
303 CCCTGGGCAAGGTTTTTCTCTATCGCTTATGTTGAGATGGGTGAGAAATATAAGCA 362
QY 3385 TATCCATGCCACATTTACGTGTGAGATTTCCGGAAGCCATTTGTTGAGAGACGGGGCAT 3444
Db |||||
363 TATCCTGCTCTCTAACAGGATCAGAGGCGGAAACCTCTTTTGGAGAAATAGGACAC 422
QY 3445 ACCATCATGATCTCTTCTGAGACTTCAGAAATACCAGTATACCTTGC 3493
Db |||||
423 ACTATTAGCACTTACTTGTGACCTTCGAAATTTACCAGTATACCTTGC 471

RESULT 14

US-09-983-965-4229/c

; Sequence 4229, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Rao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4229
; LENGTH: 437
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (335)
; OTHER INFORMATION:
; OTHER INFORMATION: Clone ID: 61-LIB3058-003-Q1-K1-H2
US-09-983-965-4229

Query Match 3.2%; Score 154; DB 9; Length 437;
Best Local Similarity 88.8%; Pred. No. 3e-31;
Matches 166; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 4064 GCACAGAGTGTCTTCTAGAAAAGATGACCAAGCTTTTGGCTCTCTCTCTCTCACC 4123
Db |||||
379 GGCAACAGGTGTCTTCTAGAGATGATGACCAACTGCTGAGTCTGAGTCTGAG 320
QY 4124 ATCAGAGTAGATTGACTGAGCATGTTGCCAAGCTTTTGGCTCTCTCTCTCTCACC 4183
Db |||||
319 ATCAGAGCGGACTAACGGAGCATGTTGCCAAGGCTTTTGGCTCTCTCTCTCACC 260
QY 4184 TGAAACTTCTGAAGGAAGATGAATGACCAACTGGGACTACGTGTGACACTGACTCAG 4243
Db |||||
259 TGAAGCTTCTGAAGGAAGATGAATGACCAACTGGGACTTCGTGTGACACTGACTCAG 200
QY 4244 ATCAGGT 4250
Db |||||
199 ATCAGGT 193

RESULT 15

US-09-764-878-357
; Sequence 357, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 357
; LENGTH: 118
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-357

Query Match 2.4%; Score 118; DB 9; Length 118;
Best Local Similarity 100.0%; Pred. No. 9.7e-22;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2058 CGATCAATGGGAACATTTAGAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCTT 2117

Db		1	CGATACAAATGGGAAACATTTAGAAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCCT	60
Qy		2118	TGCATTAGCTCCAGATAGCCACAGATAATGATCTCAGAGCTGGTCAGTTTGGAAATTCT	2175
Db		61	TGCATTAGCTCCAGATAGCCACAGATAATGATCTCAGAGCTGGTCAGTTTGGAAATTCT	118

Search completed: May 18, 2004, 14:34:34
Job time : 1253.4 secs

this Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 18:15:22 ; Search time 399.489 Seconds
(without alignments)

15102.280 Million cell updates/sec

Title: US-09-744-167-1

Perfect score: 4839

Sequence: 1 gcatactgaatcagcaggac.....atcaaaaaaaaaaaaaaaaaaa 4839

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1745496 seqs, 623391994 residues

Total number of hits satisfying chosen parameters: 3490992

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents NA New:
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	318	6.6	332	6	US-10-793-479-34465
2	160	3.3	164	6	US-10-793-479-11208
3	112.8	2.3	1888	1	PCT-US03-41761-41044
4	112.8	2.3	1888	1	PCT-US03-41761-41044
5	63.8	1.3	701	5	US-09-969-034-1233
6	53	1.1	396	6	US-10-779-543-3795
7	50	1.0	2138	6	US-10-100-683-4908
8	50	1.0	2309	6	US-10-100-683-2353
9	42.6	0.9	30612	6	US-10-724-972A-326
10	42.4	0.9	30000	1	PCT-US04-07412-382
11	41.8	0.9	403035	7	US-60-568-219-6241
12	41	0.8	1286	1	PCT-US02-39555A-2125
13	40.8	0.8	1400	7	US-60-545-213-497
14	40.8	0.8	1400	7	US-60-545-213-4769
15	40.8	0.8	6129	6	US-10-791-219-10
16	40.8	0.8	6297	6	US-10-791-219-2
17	40.8	0.8	178930	7	US-60-568-845-2909
18	40.6	0.8	605	5	US-10-779-543-20366
19	40.4	0.8	67152	7	US-60-563-440-11937
20	40.2	0.8	1790242	6	US-10-767-471-10805
21	39.4	0.8	4269	6	US-10-100-683-12071
22	39.4	0.8	17386	6	US-10-681-199-10
23	39.4	0.8	308631	7	US-60-550-051-3004
24	38.8	0.8	308631	7	US-60-568-845-2891
25	38.8	0.8	393	6	US-10-779-543-11703
26	38.8	0.8	122496	7	US-60-563-440-11884

ALIGNMENTS

RESULT 1

US-10-793-479-34465
; Sequence 34465, Application US/10793479
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/10793,479
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: US/09/513,999
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 34465
; LENGTH: 332
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-793-479-34465

Query Match 6.6%; Score 318; DB 6; Length 332;
Best Local Similarity 99.7%; Pred. No. 3.2e-77;
Matches 329; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	171	CRAGGTGTAGATGAATTTGACAAACGAAGTGAACAGTTCTCTACTTTATTGGA	230
Db	1	CAAGGTGTAGATGAATTTGACAAACGAAGTGAACAGTTCTCTACTTTATTGGA	60
QY	231	TACAAAGTGAATTAAGATTCAGATCCCTCTCTACCGGCTGTCAATTAACCCACTTT	290
Db	61	TACAAAGTGAATTAAGATTCAGATCCCTCTCTACCGGCTGTCAATTAACCCACTTT	120
QY	291	GGCC-AGTGTGAATGAATCTGAGTTCTTAATGAGTCAACACCACTGGAAGTCTCT	349
Db	121	GGCCAAAGTGTGAATGAATCTGAGTTCTTAATGAGTCAACACCACTGGAAGTCTCT	180
QY	350	CCCTGGCTCATTCAGTCCCTTGACCAAGAGGAGGATCACTGTGTAATGGACAGG	409
Db	181	CCCTGGCTCATTCAGTCCCTTGACCAAGAGGAGGATCACTGTGTAATGGACAGG	240
QY	410	ACTGTAACTTAATCCAGAGATTCACCAATGTGGATTGATGAAGTGTGTCAGAG	469
Db	241	ACTGTAACTTAATCCAGAGATTCACCAATGTGGATTGATGAAGTGTGTCAGAG	300
QY	470	ACCAGTAACTTAAGAGAACTATAGTTGGG	499

Sequence 255, Appl
Sequence 88, Appl
Sequence 42, Appl
Sequence 160, Appl
Sequence 160, Appl
Sequence 3, Appl
Sequence 4, Appl
Sequence 12317, A
Sequence 8759, Ap
Sequence 6186, Ap
Sequence 11240, A
Sequence 8887, Ap
Sequence 9284, Ap
Sequence 526, Ap
Sequence 8701, Ap
Sequence 118, Ap
Sequence 4701, Ap
Sequence 45, Appl

27 38.6 0.8 3639 7 US-60-551-161-255
C 28 38.6 0.8 3902 6 US-10-108-605A-88
C 29 38.6 0.8 3996 6 US-10-808-727-42
C 30 38.6 0.8 68196 6 US-10-417-375A-160
C 31 38.6 0.8 68196 6 US-10-417-375B-160
C 32 38.4 0.8 3885 6 US-10-377-636-3
C 33 38.4 0.8 388939 6 US-10-417-375A-4
C 34 38.4 0.8 388939 6 US-10-417-375B-4
C 35 38.2 0.8 645179 6 US-10-796-280-12317
C 36 38.2 0.8 645179 6 US-10-796-307-8759
C 37 38.2 0.8 645179 7 US-60-568-219-6186
C 38 38 0.8 201 6 US-10-796-307-11240
C 39 38 0.8 201 7 US-60-568-219-8887
C 40 38 0.8 202 6 US-10-793-479-9284
C 41 38 0.8 3811 6 US-10-834-967-526
C 42 38 0.8 380939 6 US-10-796-307-8701
C 43 37.8 0.8 705 1 PCT-US04-12047-118
C 44 37.8 0.8 772 6 US-10-779-543-4701
C 45 37.8 0.8 2232 6 US-10-808-727-45

Db 301 ACCAGTAAATTAAGAGAACTATAGTTGGG 330

RESULT 2

US-10-793-479-11208
; Sequence 11208, Application US/10793479
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/10/793,479
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: US/09/513,999
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36691
; SOFTWARE: Patent.pm
; SEQ ID NO 11208
; LENGTH: 164
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 160
; OTHER INFORMATION: r-a or g

US-10-793-479-11208

Query Match 3.3%; Score 160; DB 6; Length 164;
Best Local Similarity 98.8%; Pred. No. 7.6e-34;
Matches 160; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1982 TCTTTGGAAAGCAAAATTTAGGGGAACTCAGAACCAATGTATGAGTCCATCTTTGG 2041
DB 1 TCTTTGGAAAGCAAAATTTAGGGGAACTCAGAACCAATGTATGAGTCCATCTTTGG 60

QY 2042 GAAACATCTCTAATGTGATACAAATGGGAACTTTAGAAAGTTATGAGCTGAGATCT 2101
DB 61 GAAACATCTCTAATGTGATACAAATGGGAACTTTAGAAAGTTATGAGCTGAGATCT 120

QY 2102 CCACATAGACCATGCTTCATAGCTCCAGATGCCAGATA 2143
DB 121 CCACATAGACCATGCTTCATAGCTCCAGATGCCAGATA 162

RESULT 3

PCT-US03-41761-41044
; Sequence 41044, Application PC/TUS0341761
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: CHARTERIS, Paul
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; TITLE OF INVENTION: COMPOSITIONS, METHODS, AND SYSTEMS FOR INFERRING BOVINE BREED
; FILE REFERENCE: MM1150W0
; CURRENT APPLICATION NUMBER: PCT/US03/41761
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 41044
; LENGTH: 1888
; TYPE: DNA
; ORGANISM: Bovine 19866880507027
PCT-US03-41761-41044

Query Match 2.3%; Score 112.8; DB 1; Length 1888;
Best Local Similarity 87.9%; Pred. No. 3e-20;

Matches 123; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2756 AGACGGATATTTGCTATTTCTCTGGAGTATTAACCTCAGTTGGAAGTCTCTTGGAAAGTG 2815
DB 998 AGACGAATATTTCTCTATTTCTCTGGAAAGTATAACTCAGTCCGAAAGTCTCTTGGCAAGTG 1057

QY 2816 CAATGAATCTTATTTCTGGAAGATGGCTTCCTCCCATCTCTCATCTCCACTGGTGTAAAG 2875
DB 1058 CAATGAACCTTATTTCTGAGATGGCTTCCTCCCATCTCTCATCTCCACTGGGTTAAAG 1117

QY 2876 GAGACTATCTGTGGAAGAG 2895
DB 1118 GAGGTAGTGTAGACTTAGAG 1137

RESULT 4

PCT-US03-41761-41044
; Sequence 41044, Application PC/TUS0341761
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: CHARTERIS, Paul
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; TITLE OF INVENTION: COMPOSITIONS, METHODS, AND SYSTEMS FOR INFERRING BOVINE BREED
; FILE REFERENCE: MM1150W0
; CURRENT APPLICATION NUMBER: PCT/US03/41761
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 41044
; LENGTH: 1888
; TYPE: DNA
; ORGANISM: Bovine 19866880507027
PCT-US03-41761-41044

Query Match 2.3%; Score 112.8; DB 1; Length 1888;
Best Local Similarity 87.9%; Pred. No. 3e-20;

Matches 123; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2756 AGACGGATATTTGCTATTTCTCTGGAGTATTAACCTCAGTTGGAAGTCTCTTGGAAAGTG 2815
DB 998 AGACGAATATTTCTCTATTTCTCTGGAAAGTATAACTCAGTCCGAAAGTCTCTTGGCAAGTG 1057

QY 2816 CAATGAATCTTATTTCTGGAAGATGGCTTCCTCCCATCTCTCATCTCCACTGGTGTAAAG 2875
DB 1058 CAATGAACCTTATTTCTGAGATGGCTTCCTCCCATCTCTCATCTCCACTGGGTTAAAG 1117

QY 2876 GAGACTATCTGTGGAAGAG 2895
DB 1118 GAGGTAGTGTAGACTTAGAG 1137

RESULT 5

US-09-969-034-1233
; Sequence 1233, Application US/09969034
; GENERAL INFORMATION:
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Astle, Jon H.
; APPLICANT: Carroll, Eddie III
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dwivedi, Poornima
; APPLICANT: Molino, Gary A.
; APPLICANT: Thiagalingam, Arunthathi
; APPLICANT: Lewis, Marcia E.
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially
; FILE REFERENCE: 1657/1032
; CURRENT APPLICATION NUMBER: US/09/969,034
; CURRENT FILING DATE: 2001-10-02

```

; PRIOR APPLICATION NUMBER: 60/237,271
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 4494
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 1233
; LENGTH: 701
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 517, 620, 628, 630, 634, 650, 658, 693
; OTHER INFORMATION: n = A,T,C or G
US-09-969-034-1233

```

	Query Match	1.33;	Score 63.8;	DB 5;	Length 701;
	Best Local Similarity	66.23;	Pred. No. 5.5e-07;		
	Matches 92;	Conservative 0;	Mismatches 47;	Indels 0;	Gaps 0;
Qy	2535	ACCTCCACATGTTGATGCTACTGTGGGAGT	TTTAAAGCACCCCTCGAGCAGAAAGTGGCTCA	2594	
Dd	42	ACCTTCACGACAACTTTGCCAGTCTCAGACACTTAAACACCGAGTGT	TGAGGACTATG	101	
Qy	2595	GCCACAGACAGACGGCGAGTTTGGTTTGGTGTGGGATCTTGCCCAATGGAGAAAGTTGC	2654		
Dd	102	TTCCAAAGAACAGAGAGATGATGGTTTGCGAGATGGTATATTGCCAATGGTGAAGTTGC	161		
Qy	2655	TGATGCAGCCAAATTACA	2673		
Dd	162	AGATACAACAAAATTATCA	180		

```

RESULT 6
US-10-779-543-9795
; Sequence 9795, Application US/10779543
; GENERAL INFORMATION:
; APPLICANT: Williams et al
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
; TITLE OF INVENTION: IN CANCEROUS CELLS AND THEIR METHODS OF USE II
; FILE REFERENCE: 2300-21302
; CURRENT APPLICATION NUMBER: US/10/779,543
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 10/076,555
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 09/217,471
; PRIOR FILING DATE: 1998-12-21
; PRIOR APPLICATION NUMBER: 60/068,755
; PRIOR FILING DATE: 1997-12-23
; PRIOR APPLICATION NUMBER: 60/080,664
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/105,234
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 09/297,648
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: PCT/US99/01619
; PRIOR FILING DATE: 1999-01-28
; PRIOR APPLICATION NUMBER: 60/072,910
; PRIOR FILING DATE: 1998-01-28
; PRIOR APPLICATION NUMBER: 60/075,954
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/080,114
; PRIOR FILING DATE: 1998-03-31
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 23767
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9795
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-779-543-9795

```

```
Query Match      1.1%; Score 53; DB 6; Length 396;
Best Local Similarity 59.7%; Pred. No. 0.00038;
Matches 89; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
```

Qy	2191	ACACGCTGGGTGAGTGGCTCCAGTATGGTACCGGATTCTCAGGCTCCAAATTGCATG	2250
Db	80	AGCCCGTTGGCTGGTGGAGGACCGCATGGTCCCGACAAGGAGTGTCCGAGATGATG	139
Qy	2251	AAATGTGAAGCCCAAGTTTACATTCCAAAGAGGAGCATCACTGCAGAGCATGTGGGAAG	2310
Db	140	CAGTGTGACCCCAAGTTTGACTTCTCACAGAAAGCACCACTGTCCCGCTCGCGGAAG	199
Qy	2311	GTTTTCTGTGCTTCTGCTGTAGCCTGAA	2339
Db	200	TGTTCTTGCACAGGTGCTGCAGCCGAA	228

```

RESULT 7
US-10-100-683-4908
; Sequence 4908, Application US/10100683
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS900
; CURRENT APPLICATION NUMBER: US/10/100,683
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,564
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4908
; LENGTH: 2138
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-100-683-4908

```

	Query Match	1.0%;	Score 50;	DB 6;	Length 2388;
	Best Local Similarity	63.1%;	Pred. No. 0.0066;		
	Matches 77;	Conservative 0;	Mismatches 45;	Indels 0;	Gaps 0
Qy	2218	TGGGTACCGGATTC	TCAGGCTCCAAATTC	GATGAAATGTGAAGCCAGGTTT	TACATTACCC 2277
Db	303	TGGGTTCCAGACCAT	TATGGCATCACCTCT	TATTAAGCTGACTGTGAATTC	TGGTTGCC 362
Qy	2278	AAAAGGAGGCATCA	CTCAGCAGCATGTGGGAAGGTTTT	CTGTGCTTCCTGCTGTAGCGTG	2337
Db	363	AAACGAGACACCA	TTGCAGAAATTTGGGAATCTATTT	TGTGCTGGATGCTGCCACCTG	422
Qy	2338	AA	2339		
Db	423	AA	424		

RESULT 8
US-10-100-683-2353
; Sequence 2353, Application US/10100683
; GENERAL INFORMATION:

```

; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS900
; CURRENT APPLICATION NUMBER: US/10/100,683
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2353
; LENGTH: 2309
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (13)..(13)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (652)..(652)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (677)..(677)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-100-683-2353

Query Match
Best Local Similarity 1.0%; Score 50; DB 6; Length 2309;
Matches 77; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 2218 TGGGTACCGGATTCCTCAGGCTCCAAATTCATGTAATTCGTAAGCCAGGTTTACATTCACC 2277
Db 471 TGGGTTCAGACCATATGGCATCACACTGCTATTAACCTGCTGCTGTAATTCGTGGCC 530
QY 2278 AAAAGGAGCATCATCTCAGAGCATGTGGGAGGTTTCTGCTCTCTCTGTAGCCTG 2337
Db 531 AAAAGAGACACCATTCGAGAAATGTGGGAATGATTTTGTGCTGGATGCTGCCCTG 590
QY 2338 AA 2339
Db 591 AA 592

RESULT 9
US-10-724-972A-326
; Sequence 326, Application US/10724972A
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH03-16
; CURRENT APPLICATION NUMBER: US/10/724,972A

```

```

; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 326
; LENGTH: 30612
; TYPE: DNA
; ORGANISM: S epidermidis
; US-10-724-972A-326

Query Match
Best Local Similarity 0.9%; Score 42.6; DB 6; Length 30612;
Matches 123; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 1416 TCAGGACACTATGCTAGTAGTGGAGGACAGTGTCTGGATTGCGAGATGCAGGTCTAGA 1475
Db 26835 TGAATTTATGAATACTAATCTTATTGCAATTTGCTAGAATCGTGAATCAAGATTGA 26894
QY 1476 TTTAAAAGGAAGCTTGCATTAGTGAAGTCAAGAAATGCTGATTTCTCCACTGTTATAGACAC 1535
Db 26895 TTTAAATGATGCAATGAAATCTTTTAAGAGATGAACCTTAATAATCAAGCTTTTCTGTGCA 26954
QY 1536 ACAGCAGCAATTTATCTATCTAATGTTTGTGATTCCTATGTAATGCAGACCCAGGTGT 1595
Db 26955 AGCAAGCTCAATTTATATAATTCAGATGAAGATTTAAACCAACAATTTGACCATGCTTT 27014
QY 1596 TTTCTTTTGTTCAGAACTTTTACCCTCCAAAGAGATTCAGTAAACAGAGAAAAAGAAAT 1655
Db 27015 AGTAAATGCTCGAAAGTCTTGTGCAAAAGAAATGTTAAATTTAGATGAATCAAT 27074
QY 1656 AGAGGAAAGCAAGTCAG 1672
Db 27075 TCAGGGACTCAACCAAG 27091

RESULT 10
PCT-US04-07412-382
; Sequence 382, Application PC/TUS0407412
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-rui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yuhqing
; APPLICANT: Ghosh, Malabika
; APPLICANT: Xue, Aidong J.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Dunrui
; APPLICANT: Goodrich, Kyle W.
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wehrman, Tom
; APPLICANT: Weng, Gezhi
; APPLICANT: Wang, Zhiwei
; APPLICANT: Boyle, Bryan J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 822CIP/PCT
; CURRENT APPLICATION NUMBER: PCT/US04/07412
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: US 10/389,559
; PRIOR FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 60/365,264
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/340,187
; PRIOR FILING DATE: 2001-12-12

```

PRIOR APPLICATION NUMBER: US 10/296,115
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 10/275,027
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1920
SOFTWARE: pt_FL_genes Version 6.0
SEQ ID NO 382
LENGTH: 3000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (86)..(2008)
PCT-US04-07412-382

Query Match 0.9%; Score 42.4; DB 1; Length 3000;
Best Local Similarity 50.0%; Pred. No. 0.99;
Matches 106; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 2197 CTGGTGAGGTGGCTCCAGTATGGGTACCGGATTCAGGCTCCAAATTCATGAATGT 2256
DB 1313 CTTGGGAGAGGGCCCCCACCCTGGTGTGTGCACACGCTGATGTCATGAATGC 1372
QY 2257 GAAGCCAGGTTTACATTCACCAAAAGAGGAGCATCTGCAGAGCATCTGGAGGTTTC 2316
DB 1373 GGCTGGCACTTCTCCCTCCCTCCCTGCGGCTGATCATCTGTCAGGCTGTCAGAGTCGTG 1432
QY 2317 TGTGCTTCTGCTGTAGCTGAAATGTAAACTGTATATACATGACAGAGGAGGCTAGA 2376
DB 1433 TGCCGGAAGTGTTCGCGGACACAGTACCCGCTGAAGTACCTGAAGGACAGGATGGCAAG 1492
QY 2377 GTGTGTGTATCTGCCATTCAGTCTAATGAA 2408
DB 1493 GTCTGCGAGCGCTGCTCGGGGAGCTGAAGAA 1524

RESULT 11
US-60-568-219-6241/c
Sequence 6241, Application US/60568219
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: RESPONSIVENESS TO STATIN TREATMENT, METHODS OF DETECTION AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: C0001516
CURRENT APPLICATION NUMBER: US/60/568,219
CURRENT FILING DATE: 2004-05-06
NUMBER OF SEQ ID NOS: 28217
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6241
LENGTH: 403035
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(403035)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-60-568-219-6241

Query Match 0.9%; Score 41.8; DB 7; Length 403035;
Best Local Similarity 55.0%; Pred. No. 23;
Matches 82; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 3450 CATGAATCTTCTTGACAGCTTCAGAAATACCAGTATACCTTGCCAGTAGTTCAGGTTT 3509
DB 296925 CTTGACTGCTTAATCTTACTTCTTTGAAACCCAGTGACCTTTGTTGTTATTAATAATAA 296866
QY 3510 GGTGTTGATATGGAAGTTCGNAAACTAGCATCAAAATTCGCCAGCAACAGATCAATGA 3569
DB 296865 TGTATTTGATATTAATAATCATTTGCTAGCATCAAAATTCGCCAGCAATCTTATATCTA 296806
QY 3570 GATGATGAAGCCCATGACAAAGTCCAATG 3598
DB 296805 TGTATCTTAAGACAGAGTAAGCCAGTG 296777

RESULT 12
PCT-US02-39555A-2125/c
Sequence 2125, Application PC/TUS0239555A
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Wang, Jian-rui
APPLICANT: Wang, Zhiwei
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yunging
APPLICANT: Ghosh, Malabika
APPLICANT: Xue, Aidong J.
APPLICANT: Asundi, Vinod
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Dunrui
APPLICANT: Goodrich, Ryle W.
APPLICANT: Chen, Rui-hong
APPLICANT: Wehrman, Tom
APPLICANT: Weng, Gezhi
APPLICANT: Xu, Chongjun
APPLICANT: Boyle, Bryan J.
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 820/PCT
CURRENT APPLICATION NUMBER: PCT/US02/39555A
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: US 60/339,739
PRIOR FILING DATE: 2001-12-10
PRIOR APPLICATION NUMBER: US 10/128,558
PRIOR FILING DATE: 2002-04-22
PRIOR APPLICATION NUMBER: US 60/339,453
PRIOR FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: US 60/365,384
PRIOR FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: US 60/365,091
PRIOR FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: US 60/372,615
PRIOR FILING DATE: 2002-04-12
PRIOR APPLICATION NUMBER: US 60/376,045
PRIOR FILING DATE: 2002-04-24
PRIOR APPLICATION NUMBER: US 60/372,381
PRIOR FILING DATE: 2002-04-12
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 3134
SOFTWARE: pt_FL_genes Version 6.0
SEQ ID NO 2125
LENGTH: 1286
TYPE: DNA
ORGANISM: Homo sapiens
PCT-US02-39555A-2125

Query Match 0.8%; Score 41; DB 1; Length 1286;
Best Local Similarity 47.2%; Pred. No. 1.5;

Matches 125; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
QY 2575 CCTGAGCAGAGTGGCTCAGCCACAGAGCAGAGGCGAGTTTGGTTTGGCTGATGGGATC 2634
Db 546 CTTGACACAGCTGGCTCTCTGGGGGCTCAAGGGGGCTGGCTCTCTGGGGGTT 487
QY 2635 TTGCCAATGAGAGTGGCTGATGACGCCAAATTAACAATGAATGAACCTTCTCTGCA 2694
Db 486 CCAAGGGGGATGGTTCTTGGGCTCTCCCACTCCAGCTGTTGTAAGCCCTGTTGCT 427
QY 2695 GGAACCTGGCTGTCTACAGCAGCCAGCTCAAGCCAGTAACTACAGCTCCTTACACGCA 2754
Db 426 GCTCTCAGGCTGGATCATCTCTGCTCCCACTCCAGCTGCTCAAGCCCTGCTGCTGCT 367
QY 2755 GAGACGATATTGTTGTTATTTCTCTGGGAGTAACTCAGGTGGAAGTCTCTGTTGAAAT 2814
Db 366 CTTACAGGCTGGGTCTATCTCTGCTGCCAGTCCAGCTGCTGTAAGCCCTACTGCTGCCAGT 307
QY 2815 GCAATGAATCTTATCTCTGAAGATG 2839
Db 306 CCAGCTGCTGAAGCCCTGTAGCTG 282

RESULT 13

US-60-545-213-497
; Sequence 497, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT FILING DATE: 2004-02-18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 497
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (1395)..(1396)
; OTHER INFORMATION: n is a, c, g, or t
US-60-545-213-497

Query Match 0.8%; Score 40.8; DB 7; Length 1400;
Best Local Similarity 44.6%; Pred. No. 1.8;
Matches 108; Conservative 11; Mismatches 123; Indels 0; Gaps 0;
QY 1488 TTGCATTAGTGAAGTGAAGATGTGATTTCTCCACTGTTATAGACACACAGCAGCAAA 1547
Db 1153 TTGGTTCAGTGGCAGCATGAGTTGCCATTAAATTTGTTCTAGCTATACCAAG 1212
QY 1548 TTATCTATCTAATGGTTGATTCCTATGGAATGCAAGCCAGGTTCTTTTGTTC 1607
Db 1213 GTCTCTGTGTCAAACCTTGCCACTCTATATGCACTTTGTTACTCTTTATACAAATA 1272
QY 1608 AAGACCTTACCTCCAAAGAGATTCAGTAACAGAGAAAGAAATAGAGGAAGCA 1667
Db 1273 ATATCTAAGACTTTAAARARARARARARARARARARARARARARARARARARAR 1332
QY 1668 GTCAGATGCTCTCAATATTTATGACAGAGGAAATGAGCCACAGAGGAGTGG 1727
Db 1333 AAAAAA
QY 1728 AC 1729
Db 1393 CC 1394
RESULT 14
US-60-545-213-4769

; Sequence 4769, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT FILING DATE: 2004-02-18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4769
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (1395)..(1396)
; OTHER INFORMATION: n is a, c, g, or t
US-60-545-213-4769

Query Match 0.8%; Score 40.8; DB 7; Length 1400;
Best Local Similarity 44.6%; Pred. No. 1.8;
Matches 108; Conservative 11; Mismatches 123; Indels 0; Gaps 0;
QY 1488 TTGCATTAGTGAAGTGAAGATGTGATTTCTCCACTGTTATAGACACACAGCAGCAAA 1547
Db 1153 TTGGTTCAGTGGCAGCAATGAGTTGCCATTAAATTTGTTCTAGCTATACCAAG 1212
QY 1548 TTATCTATCTAATGGTTGATTCCTATGGAATGCAAGCCAGGTTCTTTTGTTC 1607
Db 1213 GTCTCTGTGTCAAACCTTGCCACTCTATATGCACTTTGTTACTCTTTATACAAATA 1272
QY 1608 AAGACCTTACCTCCAAAGAGATTCAGTAACAGAGAAAGAAATAGAGGAAGCA 1667
Db 1273 ATATCTAAGACTTTAAARARARARARARARARARARARARARARARARARARAR 1332
QY 1668 GTCAGATGCTCTCAATATTTATGACAGAGGAAATGAGCCACAGAGGAGTGG 1727
Db 1333 AAAAAA
QY 1728 AC 1729
Db 1393 CC 1394

RESULT 15

US-10-791-219-10
; Sequence 10, Application US/10791219
; GENERAL INFORMATION:
; APPLICANT: WEISMAN, LOIS
; TITLE OF INVENTION: INTRACELLULAR SIGNALING PATHWAYS IN DIABETIC SUBJECTS
; FILE REFERENCE: IONA:048US
; CURRENT APPLICATION NUMBER: US/10/791,219
; PRIOR FILING DATE: 2004-03-02
; PRIOR APPLICATION NUMBER: 60/452,782
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 6129
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-791-219-10
Query Match 0.8%; Score 40.8; DB 6; Length 6129;
Best Local Similarity 55.7%; Pred. No. 4.1;
Matches 78; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 2201 GTGAGTGGCTCCAGTATGGGTACCGGATTCCTAGGCTCCAAATGCAATGTAAGT 2260
Db 446 GTGACCTGAACCAATCTGGATGCCAGATGCCAATGTAAAGAGTGTCTGACTGTAGTG 505

Qy	2261	CCAGGTTTACATTACACCAAAAGGAGGCATCACTGCAGAGCATGTGGGAAGGTTTCTGTG	2320
Db	506	AGAAATTACACCTTTAGCGCGAGACACCATTTGCCGACTAATGTGGGCGAGTTTCTGCA	565
Qy	2321	CTTCCTGCTGTAGCCTGAAA	2340
Db	566	GTGCTGCTGTAATCAAGAA	585

Search completed: May 18, 2004, 13:43:32
 Job time : 405.489 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 09:54:15 ; Search time 16461.8 Seconds
(without alignments)
17461.734 Million cell updates/sec

Title: US-09-744-167-3
Perfect score: 6632
Sequence: 1 actccggccggggtagctc.....ttggaggcacattttgaagt 6632

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 1 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.1

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_ov.*
- 22: em_or.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rtd.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htg_hum.*
- 40: em_htg_mus.*
- 41: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6627.2	99.9	6632	9	AB002303	AB002303 Human mRNA
2	6167.4	93.0	6280	9	BSM805726	BSM805726 Homo sapi
3	5579.8	84.1	5841	9	HSN804398	HSN804398 Homo sapi
4	4815.2	69.6	4820	9	AF434817	AF434817 Homo sapi
5	3643.8	54.9	4499	9	BC032227	BC032227 Homo sapi
6	3418.8	51.6	3793	9	AK127003	AK127003 Homo sapi
7	3337.4	50.3	5438	10	AX122241	AX122241 Mus muscu
8	3333	50.3	5480	10	BC042669	BC042669 Mus muscu
9	2655.2	40.0	5158	9	BSM809324	BSM809324 Homo sapi
10	2454.4	37.0	2902	9	BC030808	BC030808 Homo sapi
11	2254.4	34.0	252015	9	AC008382	AC008382 Homo sapi
12	2252.8	34.0	123169	9	AC008771	AC008771 Homo sapi
13	1910.4	28.8	126052	9	AC018764	AC018764 Homo sapi
14	1834	27.7	188017	2	AC146897	AC146897 Gorilla g
15	1668.8	25.2	2622	2	AK090896	AK090896 Homo sapi
16	1197.6	18.1	204269	2	AC129932	AC129932 Mus muscu
17	1149.4	17.3	110000	2	AC118331	Continuation (3 of
18	1149.4	17.3	297979	2	AC130639	AC130639 Rattus no
19	567.4	8.6	204269	2	AC129932	AC129932 Mus muscu
20	560.6	8.5	4839	9	AF104304	AF104304 Homo sapi
21	559	8.4	4769	9	BC032680	BC032680 Homo sapi
22	556.2	8.4	1977	6	AX809313	AX809313 Sequence
23	546.6	8.2	4565	9	AF130419	AF130419 Homo sapi
24	508.4	7.7	234114	2	AC111449	AC111449 Rattus no
25	500	7.5	4823	5	AF104305	AF104305 Xenopus l
26	495.8	7.5	89973	9	AL591398	AL591398 Human DNA
27	444	6.7	455	6	AX210516	AX210516 Sequence
28	318.4	4.8	533	11	G99658	G99658 S209P6519RG
29	269	4.1	405	11	G23648	G23648 human STS W
30	263.6	4.0	1237	10	BC046341	BC046341 Mus muscu
31	244	3.7	244	11	G24283	G24283 human STS W
32	238.6	3.6	887	10	BC058416	BC058416 Mus muscu
33	201.4	3.0	1444	10	BC038255	BC038255 Mus muscu
34	195	2.9	4598	3	AF239997	AF239997 Drosophil
35	193	2.9	4651	3	DME310804	DME310804 Drosophil
36	193.4	2.9	4660	3	AY051821	AY051821 Drosophil
37	181.6	2.7	289	6	BD060611	BD060611 Secrete
38	180.8	2.7	67833	3	AC004564	AC004564 Drosophil
39	180.8	2.7	179139	3	AC099307	AC099307 Drosophil
40	180.8	2.7	313634	3	AE003454	AE003454 Drosophil
41	180.6	2.7	54186	2	AC019884	AC019884 Drosophil
42	148.4	2.2	178257	2	AC024930	AC024930 Homo sapi
43	148.4	2.2	180741	9	AC091565	AC091565 Homo sapi
44	146.8	2.2	151800	9	AC138649	AC138649 Homo sapi
45	146.2	2.2	228470	2	AC136776	AC136776 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AB002303
DEFINITION Human mRNA for KIAA0305 gene, complete cds.
ACCESSION AB002303
VERSION AB002303.1 GI:2224550
KEYWORDS KIAA0305.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Nagase, T., Ishikawa, K., Nakajima, D., Chira, M., Seki, N.,
Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.
TITLE Prediction of the coding sequences of unidentified human genes.

VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro

JOURNAL
MEDLINE
PUBMED
REFERENCE

2 (bases 1 to 6632)
Direct Submission
Osamu Ohara, Kazusa DNA Research Institute;
Submitted (28-MAR-1997) Chiba 292-0812, Japan
1532-3, Yana, Kikarazu, Chiba Tel: +81-438-52-3913
(E-mail: cdnainfo@kazusa.or.jp, Yana Tel: +81-438-52-3913)

FEATURES
source

1. 6632
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HG0042"
/sex="male"
/tissue type="brain"
/clone lib="pBluescriptII SK plus"

gene

CDS

1. 6632
249. 4868
/gene="KIAA0305"
/protein id="BAA20764.1"
/db_xref="GI:2224551"

/translation="MDSYFKAASVSDLDKLLDDPEONDEODYLDQVONAYDSNHCVS
SELASSQRTSLPQDQCVNSCASSETSYGNTSSINEKTLKGLTSIQNEKNVIGLIDL
LSSVDGTSBQIPLMGRCKPCLDLSMGNLVHATNSEEEDIKLLPDPKSNADL
LIGLDSVSTPCVSDHSDTVMSQNDTSSELQNRBIGGKELGIVDTLLSDS
YNYSGTENLKKDKIFNLESIVDFMSSALTRQSSMFHAKDLQHNKSPCGLLKQVY
LVKEVDVAVTAAELCKEEDVITQSLPKNEDLCINDNSRDNKFLPDPFQKQD
KTVIKQSAOEDSKLIDKNDVIOISSALHVSVDKPSLCLPASGSMCGSLIRSK
ARGPLPOHEKDNQVATHEIIONSVLGGEPFKENELLKQEKCKSLLOSLEG
MEDRKIDPDQVIAELDGGDTSSTVESQEGSLGTHVPESDCCGEPINTPSSNDM
DQGLDYFNIDEGAKSLIDAEPLTEQYLQNTNISKFEENVNDSKSNQNDIM
KGLDGNINITFPAEAGATGESGINII CETVDKQNTIENGSLGSKSTIPVQOGLP
TSKDEINTQLSVINSQSGVGRKQLFSLPSTRSSKDLNKPDPDTIESEFSLAD
TWPTICADISTADPQSFNSNYIDIESNSEGSSFTANEDSVPEKCEGLVLQK
QPTVWPDEANPCNVKQFTTKRHHRCACGKVCVGCNCRKCKOYLEKEARVCV
VCYETISKAQAFERWSPSTSNLSKNSHSDCTTQVPEQNTSSIPEPATLPYSALKV
PVGSLGSKSEKRVAFADGLLPNGEVADTKLSGSKRCSDESPSLSPVPMVNTVD
HSHSTVEPKNETGIDTRNEIIQSPISQVPSVEKSMNTGEGLPSTGSGFTLDDVYF
AETBEPSPTGVVNSNLPSTASIDYRLDCLDKYCNKISLLPNDSDSLPPLIVASG
EKGSPVVEHPSEHQIILLEGGFHPTVFLNANLVNKFIFYSDKYWFSTNG
LHGLQABIIILLCLPNETIPKDIIFRLPITVYKDLKQKIENLDNITFTESFLSS
KDHGGFLIPTFOKDLDSLSPNPFILGILIOKLEIPWAKVPMRLMLRLGAERYAY
PAPLTSIRGRPLGEIGHTIMLLVDLRVQVTLHNDOLLHMGKSKICKIPRK
YSDVKNVLSNNEHVISIGASPTSEADSHLVICNDGIYETQANSATGHRKVTGASF
VFNGLKTSGGFLAKSSI VEDGLMVQITPETMGLRLAUREQKDKITCGKVDVLDL
KEYVDICWDABEENKGVISVSDGISLQGFPEKIKLEADFTDEKIVKCTEVFVFL
KDQLSILSTSYQFAKIAWACSAALCPHLKLSNGMNKIGLEVSIDTDMVDFQAS
BQLLPQHYLNDLSALIPVTHGTSNSSLPLELELVFFIIEHLF"

ORIGIN

Query Match 99.9%; Score 6627.2; DB 9; Length 6632;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6629; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
1 ACTCCCGCGCGGGTAGCTCTTCACTCCTCAGCGCGAGCTGCTGTCAGATTCCCAAAAG 60
1 ACTCCCGCGCGGGTAGCTCTTCACTCCTCAGCGCGAGCTGCTGTCAGATTCCCAAAAG 60
61 CTCGCGAGGGCTGTAGGGAGGTGATCTCATTCATTAAACAGCTGTGTGTCGAGTTCCC 120
61 CTCGCGAGGGCTGTAGGGAGGTGATCTCATTCATTAAACAGCTGTGTGTCGAGTTCCC 120
121 AAATCTTTATCTATCTCAGACTTCTCTCTGATCTCCAGATTCCTTATATTCAGCTGCC 180
121 AAATCTTTATCTATCTCAGACTTCTCTCTGATCTCCAGATTCCTTATATTCAGCTGCC 180
181 TTGGATATCTCTCCAGAGATGTTCTCAAGGCATACAGAAATTAATTTCTGAATAGTCTG 240

181 TTGGATATCTCTCCAGAGATGTTCTCAAGGCATACAGAAATTAATTTCTGAATAGTCTG 240
241 CAGGTAGGATGACAGTATTATTTAAAGCAGCTGTGTCAGTACCTTGACAAAACCTCCTTGATG 300
241 CAGGTAGGATGACAGTATTATTTAAAGCAGCTGTGTCAGTACCTTGACAAAACCTCCTTGATG 300
301 ATTTGAACAGAACACAGATGAACAGATTATCTCGAGATGTACAAATGATATGAT 360
301 ATTTGAACAGAACACAGATGAACAGATTATCTCGAGATGTACAAATGATATGAT 360
361 CTAACCACTGCTCAGTTTCTTCAGAGTTGGCTTCTCAGCGAACTTTCATTTGCTCCCAA 420
361 CTAACCACTGCTCAGTTTCTTCAGAGTTGGCTTCTCAGCGAACTTTCATTTGCTCCCAA 420
421 AAGACCAAGAGTGGCTTAAATAGTTGCTCTCATCAAGAACAGCTATGAAACAAATGAGA 480
421 AAGACCAAGAGTGGCTTAAATAGTTGCTCTCATCAAGAACAGCTATGAAACAAATGAGA 480
481 GTTCCCTGAATGAAAAACACCTCAAGGGACCTTACTTCTATACAAATGAAAAAATGTAA 540
481 GTTCCCTGAATGAAAAACACCTCAAGGGACCTTACTTCTATACAAATGAAAAAATGTAA 540
541 CAGGACTTGATCTTCTTCTGATGGATGGTGTACTTTCAGATGAAATCCAGCGCTTAT 600
541 CAGGACTTGATCTTCTTCTGATGGATGGTGTACTTTCAGATGAAATCCAGCGCTTAT 600
601 ATATGGGACGATGATGAACCTATCTGTGATCTCATAGTACGATGGTAACTTGTAGTTC 660
601 ATATGGGACGATGATGAACCTATCTGTGATCTCATAGTACGATGGTAACTTGTAGTTC 660
661 ATGCAACCAATAGTGAAGAAGATATTAATAAATAATTTATCCAGATGATTTTAAAGTCTAATG 720
661 ATGCAACCAATAGTGAAGAAGATATTAATAAATAATTTATCCAGATGATTTTAAAGTCTAATG 720
721 CAGATTCTTGTGATGGATGGATTTATCTTCAGTGTGATGATCTCCCTGTGTTCTTCAA 780
721 CAGATTCTTGTGATGGATGGATTTATCTTCAGTGTGATGATCTCCCTGTGTTCTTCAA 780
781 CAGACCATGATAGTATGATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
781 CAGACCATGATAGTATGATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
841 ATAGAGAAATCCGAGGAATCAAGAAATTTGGGTATATAAAGTAGATACACACTTTTCAAGAT 900
841 ATAGAGAAATCCGAGGAATCAAGAAATTTGGGTATATAAAGTAGATACACACTTTTCAAGAT 900
901 CCTATAATTTACAGTGGAAACAGAAAAATTTAAAGATATAAAGATCTTTAATCAGTTAGAT 960
901 CCTATAATTTACAGTGGAAACAGAAAAATTTAAAGATATAAAGATCTTTAATCAGTTAGAT 960
961 CAATTGTTGATTTTAAACATGTCATCTGCTTGAATCTGCGCAAGTTTCCAAAATGTTTCAATG 1020
961 CAATTGTTGATTTTAAACATGTCATCTGCTTGAATCTGCGCAAGTTTCCAAAATGTTTCAATG 1020
1021 CCAAGACAGAGCTACACACAGAGCCAGCCATGATGATGATGATGATGATGATGATGATGATGAT 1080
1021 CCAAGACAGAGCTACACACAGAGCCAGCCATGATGATGATGATGATGATGATGATGATGATGAT 1080
1081 TAAAGAGGAGAGTAGATGTGGCAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAG 1140
1081 TAAAGAGGAGAGTAGATGTGGCAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAG 1140
1141 AGACAGAGTCTTGTGACCTGCGACCTTCCGAAAAATGAGATTTATGCTTAAATGATTCAA 1200
1141 AGACAGAGTCTTGTGACCTGCGACCTTCCGAAAAATGAGATTTATGCTTAAATGATTCAA 1200
1201 ATTCAAGAGATGAAAAATTTTCAAAATTTACCTGACTTTTCTTTTTCAGGAAGATAGACTGTGA 1260
1201 ATTCAAGAGATGAAAAATTTTCAAAATTTACCTGACTTTTCTTTTTCAGGAAGATAGACTGTGA 1260
1261 TAAACCAATCTGCAACAGAGACTCAAAAAGTTTAGACTTAAAGATTAATGATGATGATGATGATGAT 1320

Db	1261	TAAACAACTCTGCACAGAGAGACTCAAAAAGTTTAGACCTTAAGGATAATGATGTAATCC	1320
Qy	1321	AGATTCTCTTCCAGCTTTACATGTTTCCAGTAAGAGATGCGCTCCTCATTTGCTCTGTC	1380
Db	1321	AGATTCTCTTCCAGCTTTACATGTTTCCAGTAAGAGATGCGCTCCTCATTTGCTCTGTC	1380
Qy	1381	TTCTCTGCTGCTGGTCTATGTTGGATCAATTAATGAAAGTAAGACACGGGGTGAATTTT	1440
Db	1381	TTCTCTGCTGCTGGTCTATGTTGGATCAATTAATGAAAGTAAGACACGGGGTGAATTTT	1440
Qy	1441	TACCTCAGCATGAACATAAAGATAATATACAAGATGCAGTACATATACATGAAGAAATAC	1500
Db	1441	TACCTCAGCATGAACATAAAGATAATATACAAGATGCAGTACATATACATGAAGAAATAC	1500
Qy	1501	AGAACAGTGTGTTCTAGTGGGGAACCAATTCAAAGAGAATGATCTTTTGAACACAGAAA	1560
Db	1501	AGAACAGTGTGTTCTAGTGGGGAACCAATTCAAAGAGAATGATCTTTTGAACACAGAAA	1560
Qy	1561	AATGTAAGCATACTCTTCCAGTCAATTAATGAAAGGATGGAAGACAGAAAGATAGTC	1620
Db	1561	AATGTAAGCATACTCTTCCAGTCAATTAATGAAAGGATGGAAGACAGAAAGATAGTC	1620
Qy	1621	CTGACACAGACAGTAATCAGAGCTGAGTCTTTGGATGGTGACACAGTCTTACAGTTG	1680
Db	1621	CTGACACAGACAGTAATCAGAGCTGAGTCTTTGGATGGTGACACAGTCTTACAGTTG	1680
Qy	1681	TAGATCTCAAGAGGGCTTTTGGCACTCATGTCCAGAGTCTTCTGATGTTGTTGAAG	1740
Db	1681	TAGATCTCAAGAGGGCTTTTGGCACTCATGTCCAGAGTCTTCTGATGTTGTTGAAG	1740
Qy	1741	GTGTTTAAATACCTTTTCAAGCAATGATATGATGGCAAGACTTAGATTACTTTAATA	1800
Db	1741	GTGTTTAAATACCTTTTCAAGCAATGATATGATGGCAAGACTTAGATTACTTTAATA	1800
Qy	1801	TTGATGAAGCGCAAAAGTGGCCCACTAATAGTGAATGCTGAACCTGATGCTCTGA	1860
Db	1801	TTGATGAAGCGCAAAAGTGGCCCACTAATAGTGAATGCTGAACCTGATGCTCTGA	1860
Qy	1861	CAGAACAGTATCTTCAGACCACTAAACATAAAGCTTTTGAAGAAATGTAAATGACTTA	1920
Db	1861	CAGAACAGTATCTTCAGACCACTAAACATAAAGCTTTTGAAGAAATGTAAATGACTTA	1920
Qy	1921	AATCGCAAAATCAATCAGATAGATGAAGGCTTAGATGATGGAACATCAATPAATATAT	1980
Db	1921	AATCGCAAAATCAATCAGATAGATGAAGGCTTAGATGATGGAACATCAATPAATATAT	1980
Qy	1981	ATTTCATGACAGAGCAGGAGCTATTGGGGAAGTCTATGTTATATATATATTTGAA	2040
Db	1981	ATTTCATGACAGAGCAGGAGCTATTGGGGAAGTCTATGTTATATATATATTTGAA	2040
Qy	2041	CAGTTGATAACAAAATACAAATAGAAATGGCTTTCTTTAGGAGAAAAGCACTATTTC	2100
Db	2041	CAGTTGATAACAAAATACAAATAGAAATGGCTTTCTTTAGGAGAAAAGCACTATTTC	2100
Qy	2101	CAGTTCAACAGGGTTACCTACCAAGTGAAGTCTGAGATTACAAATCAATATCAGTCTCTG	2160
Db	2101	CAGTTCAACAGGGTTACCTACCAAGTGAAGTCTGAGATTACAAATCAATATCAGTCTCTG	2160
Qy	2161	ATATTACAGTCAATCTGTTGGAGGGCCAGACTTAAGCAATGTTTAGCCTTCCATCAA	2220
Db	2161	ATATTACAGTCAATCTGTTGGAGGGCCAGACTTAAGCAATGTTTAGCCTTCCATCAA	2220
Qy	2221	GAACAGGAGTTCAAGGAGCTGAATGAAGCAGATGTTCCAGATACAAATAGAAAGTGAAC	2280
Db	2221	GAACAGGAGTTCAAGGAGCTGAATGAAGCAGATGTTCCAGATACAAATAGAAAGTGAAC	2280
Qy	2281	CCAGCACAGACAGATACCGTTGTTCCAAATCACTTGTGCTATAGATTCTACAGCTGATCCAC	2340
Db	2281	CCAGCACAGACAGATACCGTTGTTCCAAATCACTTGTGCTATAGATTCTACAGCTGATCCAC	2340
Qy	2341	AGGTTAGCTTCAACTCTAAATTAATGATATAGAAAGTAAATCTGAAGGTGGAATCTAGTT	2400
Db	2341	AGGTTAGCTTCAACTCTAAATTAATGATATAGAAAGTAAATCTGAAGGTGGAATCTAGTT	2400
Qy	2401	TCGTAACCTGCAAAATGAAGATTCTGTACCTGAAAAACACTTGCAAAAGAGGCTTGGTTTTGG	2460
Db	2401	TCGTAACCTGCAAAATGAAGATTCTGTACCTGAAAAACACTTGCAAAAGAGGCTTGGTTTTGG	2460
Qy	2461	GCCAGAAAACAGCCTACTTTGGGTTCTGATTCAGAGCTCCAACTGATATGAAGTCCCAAG	2520
Db	2461	GCCAGAAAACAGCCTACTTTGGGTTCTGATTCAGAGCTCCAACTGATATGAAGTCCCAAG	2520
Qy	2521	TCAAAATTTACTTTTACCAAAAGGGGACACCAATTCGCCAGCATGTTGGGAAAGTATTTTGTG	2580
Db	2521	TCAAAATTTACTTTTACCAAAAGGGGACACCAATTCGCCAGCATGTTGGGAAAGTATTTTGTG	2580
Qy	2581	GTGTCCTGTTGTAATAGGAAGTGTAACTGGAATATCTAGAAAAAGGACAGATATGTG	2640
Db	2581	GTGTCCTGTTGTAATAGGAAGTGTAACTGGAATATCTAGAAAAAGGACAGATATGTG	2640
Qy	2641	TAGTCTGCTATGAAGTCACTTAAAGCTCAGGCAATTTGAAAGGATGATGATGATCAACTG	2700
Db	2641	TAGTCTGCTATGAAGTCACTTAAAGCTCAGGCAATTTGAAAGGATGATGATGATGATCAACTG	2700
Qy	2701	GTCTTAATCTTAAGTCTTAATCATTTCTGATGAATGATCTGTCAGGCTCCTCAGGAGA	2760
Db	2701	GTCTTAATCTTAAGTCTTAATCATTTCTGATGAATGATCTGTCAGGCTCCTCAGGAGA	2760
Qy	2761	ACCAACATCCAGTATACCTTCAACAGCAACTTTTCCAGCTCTCAGCACTTAAACAACAG	2820
Db	2761	ACCAACATCCAGTATACCTTCAACAGCAACTTTTCCAGCTCTCAGCACTTAAACAACAG	2820
Qy	2821	GTGTTGAAGGAGCTATGTTTCCAAAGAACAGAGAGATGTTTGCAGATGATATATTC	2880
Db	2821	GTGTTGAAGGAGCTATGTTTCCAAAGAACAGAGAGATGTTTGCAGATGATATATTC	2880
Qy	2881	CCAATGGTGAAGTTCAGATACAAACAAATTTATCATCTGGAAGTAAAGATGTTCTGAAG	2940
Db	2881	CCAATGGTGAAGTTCAGATACAAACAAATTTATCATCTGGAAGTAAAGATGTTCTGAAG	2940
Qy	2941	ACTTTAGTCTCTCTCACTGATGTCCTATGACAGTAAACACAGTGGATCATTTCCCATTT	3000
Db	2941	ACTTTAGTCTCTCTCACTGATGTCCTATGACAGTAAACACAGTGGATCATTTCCCATTT	3000
Qy	3001	CTACTACAGTGGAAAGCCAAACAAATGACAGGAGATATTACAGAAATGAGATAATTC	3060
Db	3001	CTACTACAGTGGAAAGCCAAACAAATGACAGGAGATATTACAGAAATGAGATAATTC	3060
Qy	3061	AGAGTCTTATTTCTCAGGTTCCATCAGTGGAAATTTCTATGAAACACAGAAATGAGG	3120
Db	3061	AGAGTCTTATTTCTCAGGTTCCATCAGTGGAAATTTCTATGAAACACAGAAATGAGG	3120
Qy	3121	GCTTACCTACTTCTGGTTCACTTACACTAGATGATGTTTTCAGAAAACCTGAAGAAC	3180
Db	3121	GCTTACCTACTTCTGGTTCACTTACACTAGATGATGTTTTCAGAAAACCTGAAGAAC	3180
Qy	3181	CATCTAGTCTACTGCTGCTTAGTTACAGCAATTTTACCTATTGCTAGTATTTTCAGATT	3240
Db	3181	CATCTAGTCTACTGCTGCTTAGTTACAGCAATTTTACCTATTGCTAGTATTTTCAGATT	3240
Qy	3241	ATAGGTTACTGCTGATATTAACAAAGTATGCTGCAATAAGATTTAGTCTTCTACCTAATG	3300
Db	3241	ATAGGTTACTGCTGATATTAACAAAGTATGCTGCAATAAGATTTAGTCTTCTACCTAATG	3300
Qy	3301	ATGAGGACAGTTTGGCCCACTTCTGGTTGCAATCTGGAGAAAAGGATCAGTGCCTGTAG	3360
Db	3301	ATGAGGACAGTTTGGCCCACTTCTGGTTGCAATCTGGAGAAAAGGATCAGTGCCTGTAG	3360
Qy	3361	TAGAGAGACATCCATCTCATGAGCAGATCAITTTTCTTTTGAAGGTGAAGGCTTTCATC	3420
Db	3361	TAGAGAGACATCCATCTCATGAGCAGATCAITTTTCTTTTGAAGGTGAAGGCTTTCATC	3420
Qy	3421	CTGTTTACATTTGTCCTAAATGCTAATCTACCTGTAATGCTCAAAATTCATATTTTATTCCT	3480
Db	3421	CTGTTTACATTTGTCCTAAATGCTAATCTACCTGTAATGCTCAAAATTCATATTTTATTCCT	3480

QY 3481 CAGACAAATATGTGACTTTTCAACCAATGGATTGCAATGCTTGGGACAGCGCAAAATTA 3540
DB 3481 CAGACAAATATGTGACTTTTCAACCAATGGATTGCAATGCTTGGGACAGCGCAAAATTA 3540
QY 3541 TTATTCCTATGCTGATGTTGGCAATGAAGATACATATTCCTAAGGACATCTTCAGACTAT 3600
DB 3541 TTATTCCTATGCTGATGTTGGCAATGAAGATACATATTCCTAAGGACATCTTCAGACTAT 3600
QY 3601 TTATCACCATATATAAGATGCTCTAAAGAAATACATAGAAAACCTTCGCAATATTA 3660
DB 3601 TTATCACCATATATAAGATGCTCTAAAGAAATACATAGAAAACCTTCGCAATATTA 3660
QY 3661 CCTTTACTGAGATTTCTCAGTAGCAGGATCAGGAGGATTCCTGTTTATTACACCTA 3720
DB 3661 CCTTTACTGAGATTTCTCAGTAGCAGGATCAGGAGGATTCCTGTTTATTACACCTA 3720
QY 3721 CTTTTTCAGAAAATTGATGATCTCTATTACCAAGTAATCCTTTTCTTTGGAATCTTA 3780
DB 3721 CTTTTTCAGAAAATTGATGATCTCTATTACCAAGTAATCCTTTTCTTTGGAATCTTA 3780
QY 3781 TCCAGAACTTGAGATTCCTGCGCAAGGTTTTCCTATGCGTTTAACTTCGAGTTGG 3840
DB 3781 TCCAGAACTTGAGATTCCTGCGCAAGGTTTTCCTATGCGTTTAACTTCGAGTTGG 3840
QY 3841 GTGCAGAAATATAAGCATATCCTGCTCTCTAAACAGCATCAGAGGCGGAAAACCTCTTT 3900
DB 3841 GTGCAGAAATATAAGCATATCCTGCTCTCTAAACAGCATCAGAGGCGGAAAACCTCTTT 3900
QY 3901 TTGAGAAATAGGACACACTATTATGAACCTTACTTGTGACCTTCGAAATACAGTATA 3960
DB 3901 TTGAGAAATAGGACACACTATTATGAACCTTACTTGTGACCTTCGAAATACAGTATA 3960
QY 3961 CTTTGCATAATATAGTCAACTGTTGATTCATATGGAAATGGGAAAAGCTGCATAAAA 4020
DB 3961 CTTTGCATAATATAGTCAACTGTTGATTCATATGGAAATGGGAAAAGCTGCATAAAA 4020
QY 4021 TACCACGAAAAGTACAGTGATGTAATGAAAGTACTAAATTTCTCCAAATGAGCATGCA 4080
DB 4021 TACCACGAAAAGTACAGTGATGTAATGAAAGTACTAAATTTCTCCAAATGAGCATGCA 4080
QY 4081 TTACATTTGGCAAGTTTACAGTACAGAGCATCTCATCTAGTCTGTATGATACAGATG 4140
DB 4081 TTACATTTGGCAAGTTTACAGTACAGAGCATCTCATCTAGTCTGTATGATACAGATG 4140
QY 4141 ATGGAATTTATGAACACAGCCCAACAGTGCCTTCCATGCAATGCAAGGTTG 4200
DB 4141 ATGGAATTTATGAACACAGCCCAACAGTGCCTTCCATGCAATGCAAGGTTG 4200
QY 4201 CAAGTTTGTGATTCATGAGCTCTTAAACATCTTCAGGATTTCTTGTAACTCCA 4260
DB 4201 CAAGTTTGTGATTCATGAGCTCTTAAACATCTTCAGGATTTCTTGTAACTCCA 4260
QY 4261 GCATAGTTGAAGATGGCTTAAATGTAACAAATACTCCAGAGACCATGAATGGCTTGGCG 4320
DB 4261 GCATAGTTGAAGATGGCTTAAATGTAACAAATACTCCAGAGACCATGAATGGCTTGGCG 4320
QY 4321 TAGCTTTACGAGACAGAAAGCTTTTAAATTAATGATGGGAAAGTTGATGCGATAGACC 4380
DB 4321 TAGCTTTACGAGACAGAAAGCTTTTAAATTAATGATGGGAAAGTTGATGCGATAGACC 4380
QY 4381 TGAGAGAAATACGTTGGATATCTGCTGGGTAGATGCTGAAGAAAAGGAAACAGAGGATTA 4440
DB 4381 TGAGAGAAATACGTTGGATATCTGCTGGGTAGATGCTGAAGAAAAGGAAACAGAGGATTA 4440
QY 4441 TCAGTTCAGTGGATGGAATATCATTAAGGATTTCCAAAGTGAAGAAATATAAACTGGAAG 4500
DB 4441 TCAGTTCAGTGGATGGAATATCATTAAGGATTTCCAAAGTGAAGAAATATAAACTGGAAG 4500
QY 4501 CAGATTTGAAACCGATGAGAGATTTGTAAGTGTACCGAGGTTCTCTACTTTCTAAAGG 4560
DB 4501 CAGATTTGAAACCGATGAGAGATTTGTAAGTGTACCGAGGTTCTCTACTTTCTAAAGG 4560
QY 4561 ACCAGGATTTATCTATTTTATCAACTCTTTATAGTTTGGCAAAAGAAATAGCCATGGCTT 4620

DB 4561 ACCAGGATTTATCTATTTTATCAACTCTTTATAGTTTGGCAAAAGAAATAGCCATGGCTT 4620
QY 4621 GTAGTGTGCGTGTGCGCTCACCTGAAAACTCTAAAAAGTAATGGGATGAATAAAATTG 4680
DB 4621 GTAGTGTGCGTGTGCGCTCACCTGAAAACTCTAAAAAGTAATGGGATGAATAAAATTG 4680
QY 4681 GACTCAGAGTTTCCATTGACACTGATATGGTTGAATTTTCAGGCAGGATCTCGAAGCCCAAC 4740
DB 4681 GACTCAGAGTTTCCATTGACACTGATATGGTTGAATTTTCAGGCAGGATCTCGAAGCCCAAC 4740
QY 4741 TTTCTGCTCAGCATTTATCTAAATGATCTTGATAGTCTCTGATACCTGTGATCCATGGTG 4800
DB 4741 TTTCTGCTCAGCATTTATCTAAATGATCTTGATAGTCTCTGATACCTGTGATCCATGGTG 4800
QY 4801 GGACCTCCAACTCTAGTTTACCATTTAGAAATAGAAATAGTCTTTTCAATATAGAACATC 4860
DB 4801 GGACCTCCAACTCTAGTTTACCATTTAGAAATAGAAATAGTCTTTTCAATATAGAACATC 4860
QY 4861 TTTTGTAGTGAAGAAATGTGCAATATTACATATTGCAACCTAAATTTGTTAAACCTAAC 4920
DB 4861 TTTTGTAGTGAAGAAATGTGCAATATTACATATTGCAACCTAAATTTGTTAAACCTAAC 4920
QY 4921 CAGCACTAAAGCTGAAATGCCAACAACTAAAGTATAATATGCTCTGATTTTGAAC 4980
DB 4921 CAGCACTAAAGCTGAAATGCCAACAACTAAAGTATAATATGCTCTGATTTTGAAC 4980
QY 4981 ACATAAGCTTGTCTCTTTAGGAGCAATGATCTTTTCAATCATTTAGCACAAATATTAAA 5040
DB 4981 ACATAAGCTTGTCTCTTTAGGAGCAATGATCTTTTCAATCATTTAGCACAAATATTAAA 5040
QY 5041 TATCTAAATAATTAAGAGATCATACTTTCTGTAGCTTTTCAATTAATTAAGTACTAAA 5100
DB 5041 TATCTAAATAATTAAGAGATCATACTTTCTGTAGCTTTTCAATTAATTAAGTACTAAA 5100
QY 5101 AAGCAAGATTTCTTTTAAAGAAATTTATAGCAATTTACTGTGTTATTTAAATGCTAAGCC 5160
DB 5101 AAGCAAGATTTCTTTTAAAGAAATTTATAGCAATTTACTGTGTTATTTAAATGCTAAGCC 5160
QY 5161 AAGTATCTGCACCTTTAGGTATACCTCTTTATGCCAATAATGATTTTAATGAAGCTCTTT 5220
DB 5161 AAGTATCTGCACCTTTAGGTATACCTCTTTATGCCAATAATGATTTTAATGAAGCTCTTT 5220
QY 5221 TCAGATGTAACCTTATGAGGAATATCTGTTGTGTATGTCAGTGTAGAACTAGCTGGT 5280
DB 5221 TCAGATGTAACCTTATGAGGAATATCTGTTGTGTATGTCAGTGTAGAACTAGCTGGT 5280
QY 5281 TTTCTAAAGTCTGTCAAATTTGATTTTTCAGTGGCACAACCAACAGTTTTCAGGTCTTAGACT 5340
DB 5281 TTTCTAAAGTCTGTCAAATTTGATTTTTCAGTGGCACAACCAACAGTTTTCAGGTCTTAGACT 5340
QY 5341 TATAATTTCTTGAATPAAACCTGATTAATTTTGTATTAATTTGGAGTGGAGACCTACCTCC 5400
DB 5341 TATAATTTCTTGAATPAAACCTGATTAATTTTGTATTAATTTGGAGTGGAGACCTACCTCC 5400
QY 5401 ATAATTTAGTAAACCTCTTTTGGATTATAACAGAAATTTTCCCTTTTCTTCTCAAAT 5460
DB 5401 ATAATTTAGTAAACCTCTTTTGGATTATAACAGAAATTTTCCCTTTTCTTCTCAAAT 5460
QY 5461 ATTACATATGATGATTTATATATCCATATATAGTTTTCCTGATTAATTAATGATATTA 5520
DB 5461 ATTACATATGATGATTTATATATCCATATATAGTTTTCCTGATTAATTAATGATATTA 5520
QY 5521 AAATAATTTGCGGTGCTTCAGACCTTTTGTCTCTATATTTAAAGTATATTTGTTTTATAG 5580
DB 5521 AAATAATTTGCGGTGCTTCAGACCTTTTGTCTCTATATTTAAAGTATATTTGTTTTATAG 5580
QY 5581 CAAGAACATATTTCTGAATGTTTTTAAATCTTTTAAATTTTATATGATAGGTAATTTTT 5640
DB 5581 CAAGAACATATTTCTGAATGTTTTTAAATCTTTTAAATTTTATATGATAGGTAATTTTT 5640
QY 5641 GTATCACAATGCAATTTATTTTCTCTCTCTTCCAACTATACACCTGATTTTACC 5700

Db	5641	GTATCAATGCAATATTTTTCCTCCTTCTTCCAACTATACCACTGATTTACC	5700
Qy	5701	ACTTCTAAGAGTACTGACGACGGCCAGATGACCCCTTGAAGTAGTCATTTATGATGCAAT	5760
Db	5701	ACTTCTAAGAGTACTGACGACGGCCAGATGACCCCTTGAAGTAGTCATTTATGATGCAAT	5760
Qy	5761	AAATGAAGCCTGAAACAGGTTTTTTTACTTCCACTTTAAATCCTTTAGAAATTTCTTGGCAA	5820
Db	5761	AAATGAAGCCTGAAACAGGTTTTTTTACTTCCACTTTAAATCCTTTAGAAATTTCTTGGCAA	5820
Qy	5821	CTTCGCATATTTTCATGACACTGGTGATATAGTATATAAATTTAAATGAACATAATCTTT	5880
Db	5821	CTTCGCATATTTTCATGACACTGGTGATATAGTATATAAATTTAAATGAACATAATCTTT	5880
Qy	5881	TGCATATTTTAAATCTTTTATATAGTATATTTTATACAGGATATTAACATAGTT	5940
Db	5881	TGCATATTTTAAATCTTTTATATAGTATATTTTATACAGGATATTAACATAGTT	5940
Qy	5941	AAATCCCTATGTTTGAATTTGTACAGAGCTTTCTCTTACTTCAAAACAGCAAAAAG	6000
Db	5941	AAATCCCTATGTTTGAATTTGTACAGAGCTTTCTCTTACTTCAAAACAGCAAAAAG	6000
Qy	6001	TGGGGGCGCATATGCTAGTCTGCTGATTAAGTATGTAAGGATTTAAATCAATTTTGA	6060
Db	6001	TGGGGGCGCATATGCTAGTCTGCTGATTAAGTATGTAAGGATTTAAATCAATTTTGA	6060
Qy	6061	TGCTTTAAACATCTCATGTGATATATGTTTGTATCAAAACACTCATATTTTCA	6120
Db	6061	TGCTTTAAACATCTCATGTGATATATGTTTGTATCAAAACACTCATATTTTCA	6120
Qy	6121	AGAAAAAGAAATATGTTAAATAGCCCTGTTTAAAGAAAATATTTATGAAGCATCTCAA	6180
Db	6121	AGAAAAAGAAATATGTTAAATAGCCCTGTTTAAAGAAAATATTTATGAAGCATCTCAA	6180
Qy	6181	CTTGACATCAAGTCAAGTATTAACCTCAGACTGAGGCTCAAGCTAGGAGACTGA	6240
Db	6181	CTTGACATCAAGTCAAGTATTAACCTCAGACTGAGGCTCAAGCTAGGAGACTGA	6240
Qy	6241	GAATTTTAAATCAGTTTCGGGCATATAGTTTGGCATGATCATCTGTAGTACTTAGCCAA	6300
Db	6241	GAATTTTAAATCAGTTTCGGGCATATAGTTTGGCATGATCATCTGTAGTACTTAGCCAA	6300
Qy	6301	AGACAAATTTGGAGAGATATACAGCTTCTGGAAGTAGTACTTCTGGAACATGTAAG	6360
Db	6301	AGACAAATTTGGAGAGATATACAGCTTCTGGAAGTAGTACTTCTGGAACATGTAAG	6360
Qy	6361	TGTCGAGATATCAATAAATGGAACCTGTTTATAATTTTGTGAAATTTATTCGAAATGGT	6420
Db	6361	TGTCGAGATATCAATAAATGGAACCTGTTTATAATTTTGTGAAATTTATTCGAAATGGT	6420
Qy	6421	GTAAGATGAAACAAATTCGATATCAAAACCCCAATTTATGTTTCTTAATATAGTATGTA	6480
Db	6421	GTAAGATGAAACAAATTCGATATCAAAACCCCAATTTATGTTTCTTAATATAGTATGTA	6480
Qy	6481	TTCTGCCATGTAAGTATTAAGCACTGTTAAATTAACCAATGATAGAGGCTGTTCCAT	6540
Db	6481	TTCTGCCATGTAAGTATTAAGCACTGTTAAATTAACCAATGATAGAGGCTGTTCCAT	6540
Qy	6541	GATGGGACAGCTTTGGATTTGTTTTCATAAATCTTACATTTCAATAAATTTGGAATTA	6600
Db	6541	GATGGGACAGCTTTGGATTTGTTTTCATAAATCTTACATTTCAATAAATTTGGAATTA	6600
Qy	6601	TGTCCTCGAAGTTTGGAGGCAATTTTGAAGT	6632
Db	6601	TGTCCTCGAAGTTTGGAGGCAATTTTGAAGT	6632

RESULT 2

HSM805726

LOCUS

DEFINITION

Homo sapiens mRNA; cdna DKFZp686K236 (from clone DKFZp686K236);

complete cds.

BX537424

ACCESSION

HSM805726

LOCUS

DEFINITION

Homo sapiens mRNA; cdna DKFZp686K236 (from clone DKFZp686K236);

complete cds.

BX537424

ACCESSION

BX537424.1 GI:31873349

Homo sapiens (human)

Homo sapiens

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 6280)

Ansoer, H.W., Krieger, S., Regiert, T., Rittmüller, C., Schwager, B.,

Wiemann, S., Weill, B., Amid, C., Osanger, A., Fobo, G., Han, M. and

Wiemann, S.

Direct Submission

Submitted (16-JUN-2003) M2FS, Ingolstaedter Landstr.1, D-85764

Neuberberg, S.

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKZ); Email: s.wiemann@dkfz-heidelberg.de;

sequenced by EMBL (European Molecular Biology Laboratories,

Heidelberg/Germany) within the cDNA sequencing consortium of the

German Genome Project.

This clone (DKFZp686K236) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further

information about the clone and the sequencing project is available

at http://mips.gsf.de/proj/cDNA/.

Location/Qualifiers

1..6280

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/map="5q14.1"

/clone="DKFZp686K236"

/tissue_type="human cervix"

/clone_lib="686 (synonym: hicc3). Vector pSport1_Sfi; host

DH10B; sites SfiIA + SfiIB"

/dev_stage="adult"

1..6280

/gene="DKFZp686K236"

133..4752

/gene="DKFZp686K236"

/note="endofin"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAD97666.1"

/db_xref="GI:31873350"

/translation="MDSYFKAASVSLDKLLDDFEONPEQDYLODYQVAYDSNHCVS

SELASQKTSILPKDQECVNCASSETSYGNESLNKTKLGLTSIQEKVNTGLDL

LSSVPGGVSDEIPLYMRCRSPKCDLIDMGNLVHATNSEEIDKLLPDDFKNADS

LIGLDSVSTPCVSTSDHSDTVREQNDTSSELRNREIGIKELGKVDITLSDS

YNSSTENLKKIKFNQLESIVDFNMSALTQSSKMFHAKLQHKQPCQLKLDVG

LVKEVDVAVITAAELKEGKTSALTCTLPKNECLDNDNSRDNFKLPDFSFQED

KTVIKSAQEDSKSLDKNDVIODSSALHVSCKDVPSSLCLPASGMSGLTESK

AKGDLPOHEKNDIQDVTIHEEQNSVVLGGEPKENDLLKQKCKSILLOSLEG

MEDRKIDPDQTVIRAESLUGGDTSTVVESEGLSGTHVPSSDCCGFINTFSSNDM

DQDDLYFNIDEGAKSGPLIDAELEDAFTLQYLTNTIKSFENNDSKQMNQIDM

KGLDDGNINNTVFAEAGAIGESHGINICETVDKONTIENGSLGKSTIPVOOGLP

TSKSRITNOLSDINSQSVGARGKQLFSLPSTRSSKDLNKPDPVPTIIESEPTAD

TVPTTCALDSTADPQVSFNSYIDIESNSGSGSFVTANEDSVENCKEGLVGOK

QPTWVDEANCMNCOVKFTFKERHCRAGCKVFCVCCNCKCKLOYLEKEARVCV

VCETILSKAFAERMVSTGNLKSNDDECTTVQPPENQTSSTSPSPATLPSALKQ

POVEGCKEQRVMPFADLPNGEVADTKLSSGSKRCSDFSPSLDPMVNTVID

HSHTTVKPNNETGDIITNEIISQIPVSEKLSNMKNIGNEGLTSGSFTLDDDF

ABTEPSSPTGLVNSNPLFIASISDRIILCKINKYCNKISLLPNDESLPLLVASG

BKSGVPVBEHPHQHEQITLLEGEFHPVFLVNLNVLNVKFIYSSDKWFTSTNG

LHGLQAEIILLLCLPNEDTIPKDIPLFITIKDALKGVKIEIENLNTIYFESLSS

KDHGSLFTPTFKLDDLSLPNPLGLIQLKLEIPKVPFNLMLRGAAYKAY

PAPLTSIRGRKFLFGEIHTINLLVLDNRYQVTLHNIDQLLINEMKSKCIKPRK

YSDVMKLNNSNEHVISIGASFSTEADSHLVCIQNDGIYFTQANSTIGPRKVTGASF

VYFGALKTSSGFLAKSVIVEDGLMVQITPTMNGRLALREQDFKTCGKVTADVL

REYVDICWDAAEKGKNGVISVDGILSQGFPSEKIKLEADFETDEKIVCTEYFYL

KQDQLSILTSYQFAKEIAMCAALCPHLTKLKSNGMKICLRVSIIDTDMVEFOAGS

EQQLLPQHYLNDLDSALLPVIHVGTSNSSLPLEIELVFFIEHLF"

Query Match

93.0%; Score 6167.4; DB 9; Length 6280;

ORIGIN

Best Local Similarity 99.6%; Pred. No. 0;				Matches 6191; Conservative 0; Mismatches 21; Indels 1; Gaps 1;			
QY	185	ATATCTCTCCAGGATGTTCTCAAGGCATACAAGAAATTAATTTCTGAATAGTCTGCAGG	244	Db	1089	AAGAGATGAAATTTTCAAATTTACCTGACCTTTTCCTTTTTCAGGAAGATAAGACTGTTTATAA	1148
Db	69	AAAGCTCCGCGGGGCTGTAGGGAGGCATACAAGAAATTTAAATTTCTGAATAGTCTGCAGG	128	QY	1265	ACAATCTGCACAAGAGACTCAAAAAGTTTACACCTTAAGCAATATGATGTAATCCAAGA	1324
QY	245	TAGGATGGACAGTTATTTTAAAGACAGCTGTCTGAGTCACTTGCACAAATCTCTTGATGATTT	304	Db	1149	ACATCTGCACAAGAGACTCAAAAAGTTTACACCTTAAGCAATATGATGTAATCCAAGA	1208
Db	129	TAGGATGGACAGTTATTTTAAAGACAGCTGTCTGAGTCACTTGCACAAATCTCTTGATGATTT	188	QY	1325	TTCTCTTTCAGCTTTACATGTTTCCAGTAAAGATGTCGCGTCTCATTTGCTGCTCTTCC	1384
QY	305	TGAACAGAACCCAGATGAACAGATTTATCTGCGAGATGTACAAAATGCAATATGATTTCTAA	364	Db	1209	TTCTCTTTCAGCTTTACATGTTTCCAGTAAAGATGTCGCGTCTCATTTGCTGCTCTTCC	1268
Db	189	TGAACAGAACCCAGATGAACAGATTTATCTCAAGATGTACAAAATGCAATATGATTTCTAA	248	QY	1385	TGCGTCTGGCTCTATCTGTGGATCATTTAATTCGAAAGTAAAGCACGGGCTGATTTTTTACC	1444
QY	365	CCACTGCTCAGTTCTTCCAGAGTTGGCTTCTCTCAGCGAACTTTCATTTGCTCCCAAGA	424	Db	1269	TGCGTCTGGCTCTATCTGAGTGGATCATTTAATTCGAAAGTAAAGCACGGGCTGATTTTTTACC	1328
Db	249	CCACTGCTCAGTTCTTCCAGAGTTGGCTTCTCTCAGCGAACTTTCATTTGCTCCCAAGA	308	QY	1445	TCAGCATGAACATAAAGATAATATACAAGATGCACTGACTATACATGAAGAAATACAGAA	1504
QY	425	CAAAGTGGCTTAATAGTTTGGCTCATCGAAACCAAGCTTATCGAAACAAATGAGTTTC	484	Db	1329	TCAGCATGAACATAAAGATAATATACAAGATGCACTGACTATACATGAAGAAATACAGAA	1388
Db	309	CCAAGTGGCTTAATAGTTTGGCTCATCGAAACCAAGCTTATCGAAACAAATGAGTTTC	368	QY	1505	CAGTGTGTTCTAGTGGGAACCAATTCAGAGAGATGATCTTTTGAACAGGAATATG	1564
QY	485	CCTGAATGAAAAACACTCAAGGGACTTACTTCTATACAAATGAAAAATGTAACAGG	544	Db	1389	CAGTGTGTTCTAGTGGGAACCAATTCAGAGAGATGATCTTTTGAACAGGAATATG	1448
Db	369	CCTGAATGAAAAACACTCAAGGGACTTACTTCTATACAAATGAAAAATGTAACAGG	428	QY	1565	TAAAGCATACTCTCTTCACTCATTAATTAAGGGATGAAGACACAGAAAGATAGATCTCTGA	1624
QY	545	ACTTGATCTTCTTCTCTGCTGGATGGTGTACTTTCAGATGAAATCCAGCGCTTATATAT	604	Db	1449	TAAAGCATACTCTCTTCACTCATTAATTAAGGGATGAAGACACAGAAAGATAGATCTCTGA	1508
Db	429	ACTTGATCTTCTTCTCTGCTGGATGGTGTACTTTCAGATGAAATCCAGCGCTTATATAT	488	QY	1625	CCAGACAGTAATCAGAGCTGAGTCTTTTGGATGGTGGTGGACACAGCTTCTACAGTTGTAGA	1684
QY	605	GGGACGATGTAGTAAACCTATCTGTGATCTGATTAAGTGAATGATGTTTAACTAGTTCATGC	664	Db	1509	CCAGACAGTAATCAGAGCTGAGTCTTTTGGATGGTGGTGGACACAGCTTCTACAGTTGTAGA	1568
Db	489	GGGACGATGTAGTAAACCTATCTGTGATCTGATTAAGTGAATGATGTTTAACTAGTTCATGC	548	QY	1685	ATCTCAAGAGGGGCTTCTGGCACCTCATGTCCACAGCTTCTCTGATCTTGTGAAGGTTT	1744
QY	665	AACCAATAGTGAAGAGATATTAATAATTTATGCCAGATGATTTAAGTCTAATGCGA	724	Db	1569	ATCTCAAGAGGGGCTTCTGGCACCTCATGTCCACAGCTTCTCTGATCTTGTGAAGGTTT	1628
Db	549	AACCAATAGTGAAGAGATATTAATAATTTATGCCAGATGATTTAAGTCTAATGCGA	608	QY	1745	TATTAATACTTTTTCAAGCAATGATATGGATGGGCAAGACTTAGATTAATTTAATTTGA	1804
QY	725	TTCTTGATGGATGGATTTATCTTCACTGTCAGATCTCCCTGCTGTTTCTTCAACAGA	784	Db	1629	TATTAATACTTTTTCAAGCAATGATATGGATGGGCAAGACTTAGATTAATTTAATTTGA	1688
Db	609	TTCTTGATGGATGGATTTATCTTCACTGTCAGATCTCCCTGCTGTTTCTTCAACAGA	668	QY	1805	TGAAGCGCAAAAAGTGGCCCACTAATTAGTGAATGCTGAATGCTGCTTCTTGACAGA	1864
QY	785	CCATGATAGTACTGTCCAGAACCAACAGATGATAGTCTGCAATTTACAAAATAG	844	Db	1689	TGAAGCGCAAAAAGTGGCCCACTAATTAGTGAATGCTGAATGCTGCTTCTTGACAGA	1748
Db	669	CCATGATAGTACTGTCCAGAACCAACAGATGATAGTCTGCAATTTACAAAATAG	728	QY	1865	ACAGTATCTTCAGACCACTTAAATAAAGTCTTTTGAAGAAAATGTAATGACTCTAAATC	1924
QY	845	AGAAATCGGAGGAATCAAAGAAATGGGTATATAAAGTAGATACACACTTTCAGATTCCTA	904	Db	1749	ACAGTATCTTCAGACCACTTAAATAAAGTCTTTTGAAGAAAATGTAATGACTCTAAATC	1808
Db	729	AGAAATCGGAGGAATCAAAGAAATGGGTATATAAAGTAGATACACACTTTCAGATTCCTA	788	QY	1925	GCAATGATCAGATAGATATGAAAGGCTTAGATGATGAAACATCAATAATATATATTT	1984
QY	905	TAAATACAGTGAACAGAAATTTAAAGATTAAGATTAAGATCTTTAATCAGTTAGATCAAT	964	Db	1809	GCAATGATCAGATAGATATGAAAGGCTTAGATGATGAAACATCAATAATATATATTT	1868
Db	789	TAAATACAGTGAACAGAAATTTAAAGATTAAGATTAAGATCTTTAATCAGTTAGATCAAT	848	QY	1985	CAATGCAGAACGAGGCTTATTTGGGAAAAGTCACTGGTATTAATATAATTTTGTGAACAGT	2044
QY	965	TGTTGATTTTAACTGTCATCTGTTTGAATCGCAAAAGTTTCCAAAATGTTTTCATGCCAA	1024	Db	1869	CAATGCAGAACGAGGCTTATTTGGGAAAAGTCACTGGTATTAATATAATTTTGTGAACAGT	1928
Db	849	TGTTGATTTTAACTGTCATCTGTTTGAATCGCAAAAGTTTCCAAAATGTTTTCATGCCAA	908	QY	2045	TCATAAACAATAAATAAATAGAAAATGCGCTTTCTTTAGGAGAAAAGACACTATTCAGT	2104
QY	1025	AGACAGCTACACACAGAGCCGACCTGAGATTAAGATGTTTGGCTTAGTAAA	1084	Db	1929	TCATAAACAATAAATAAATAGAAAATGCGCTTTCTTTAGGAGAAAAGACACTATTCAGT	1988
Db	909	AGACAGCTACACACAGAGCCGACCTGAGATTAAGATGTTTGGCTTAGTAAA	968	QY	2105	TCACAAAGGGTTTACCTACCAGTAACTGAGATTAACAATCAATTTATCAGTCTCTGATAT	2164
QY	1085	AGAGAACTAGATGTGGCAGTCAATAACTCGCGCAGAAATGTTTAAAGAAAGAGGGCAGAC	1144	Db	1989	TCACAAAGGGTTTACCTACCAGTAACTGAGATTAACAATCAATTTATCAGTCTCTGATAT	2048
Db	969	AGAGAACTAGATGTGGCAGTCAATACTGCGCAGAAATGTTTAAAGAAAGAGGGCAGAC	1028	QY	2165	TAAACAGTCAATCTGTTGGAGGGCCAGACCTTAAGCAATTTGTTTACGCTTCCATCAAGAAC	2224
QY	1145	AAGTCTTTGACCTGACGCTTCCGAAAATGAAAGATTTATGCTTAAATGATTTCAAAATTC	1204	Db	2049	TAAACAGTCAATCTGTTGGAGGGCCAGACCTTAAGCAATTTGTTTACGCTTCCATCAAGAAC	2108
Db	1029	AAGTCTTTGACCTGACGCTTCCGAAAATGAAAGATTTATGCTTAAATGATTTCAAAATTC	1068	QY	2225	AGGAGTTCAAAGGACCTTGAATAGCCAGATGTTCCAGATCAATAGAAAGTGAACCCAG	2284
QY	1205	AAGAGATGAAATTTCAAATTTACCTGACTTTTCTCTTTCAGGAAGATAGACTGTTTATAA	1264	Db	2109	AGGAGTTCAAAGGACCTTGAATAGCCAGATGTTCCAGATCAATAGAAAGTGAACCCAG	2168
Db				QY	2285	CACAGCAGATACCGTGTGTTTCCCAATCACTGTTGCTATAGATTTCTACAGCTGATCCACAGGT	2344
				Db	2169	CACAGCAGATACCGTGTGTTTCCCAATCACITTTGCTATAGATTTCTACAGCTGATCCACAGGT	2228

4389 TTTTGAACCGAATGAGAAATGTAATAATGTACCGAGGTGTTCTTCTTAAAGGACCA 4448
4565 GGATTTATCTATTTTATCAACTCTTTATCAGTTTGCACAAAAGAAATAGCCATGGTTGTAG 4624
4449 GGATTTATCTATTTTATCAACTCTTTATCAGTTTGCACAAAAGAAATAGCCATGGTTGTAG 4508
4625 TCCTGGCTGTCGCTCCTCCTGAAACTCTTAAAGTAAATGGGATGAATAAATGGACT 4684
4509 TCCTGGCTGTCGCTCCTCCTGAAACTCTTAAAGTAAATGGGATGAATAAATGGACT 4568
4685 CAGAGTTTCCATTGACACTGATATGGTTGAAATTCAGGACGAGATCTGAAGGCCCAACTTCT 4744
4569 CAGAGTTTCCATTGACACTGATATGGTTGAAATTCAGGACGAGATCTGAAGGCCCAACTTCT 4628
4745 GCCTCAGCATATCTAAATGATCTTTGATGATGCTCTGATACCTGTGATCCATGGTGGAC 4804
4629 GCCTCAGCATATCTAAATGATCTTTGATGATGCTCTGATACCTGTGATCCATGGTGGAC 4688
4805 CTCCAACTCTAGTTTACCAATAGAAATAGAAATAGTGTGTTTTTCATTATAGAACATCTTTT 4864
4689 CTCCAACTCTAGTTTACCAATAGAAATAGAAATAGTGTGTTTTTCATTATAGAACATCTTTT 4748
4865 TTAGTGAAGAATGTGCCATATACATATGCAACCTAAATTTGTTTAAACTTAACCTCCAGC 4924
4749 TTAGTGAAGAATGTGCCATATACATATGCAACCTAAATTTGTTTAAACTTAACCTCCAGC 4808
4925 ACTAAGCTGAAATGCCCAACACTAAAGTATTAATATGCTGATTTTGAACACAT 4984
4809 ACTAAGCTGAAATGCCCAACACTAAAGTATTAATATGCTGATTTTGAACACAT 4868
4985 AAGCTTTGCTCTTTAGGACGAAATGATCTTTTCAAACTAATAGCACAAATTTTAAATATC 5044
4869 AAGCTTTGCTCTTTAGGACGAAATGATCTTTTCAAACTAATAGCACAAATTTTAAATATC 4928
5045 TAAATTTTAAAGATCCATCTTTCTAGCTTTTACAAATTTTAAAGTAAATTTTAAAGTAAAG 5104
4929 TAAATTTTAAAGATCCATCTTTCTAGCTTTTACAAATTTTAAAGTAAATTTTAAAGTAAAG 4988
5105 CAAGATTTCTTTTAAAGATTTTATAGCAATTTTCTGTTTATTTTAAATGCTTAAGCCAAAG 5164
4989 CAAGATTTCTTTTAAAGATTTTATAGCAATTTTCTGTTTATTTTAAATGCTTAAGCCAAAG 5048
5165 TATCTGCATTAAGTATACCTCTTTATGCCAATAATGATTTTAAATGAGGCTCTTTTTCAG 5224
5049 TATCTGCATTAAGTATACCTCTTTATGCCAATAATGATTTTAAATGAGGCTCTTTTTCAG 5108
5225 ATGTAACTTATGAGGAAATATCTGCTTTGTTATATGCCAGTTAGAACTAGTCTGTTTCT 5284
5109 ATGTAACTTATGAGGAAATATCTGCTTTGTTATATGCCAGTTAGAACTAGTCTGTTTCT 5168
5285 AAAGTCTCTCAAAATTTTATTTTCAAGTGCACAAAACCAAGTTTTCCTCAAAATTTATA 5344
5169 AAAGTCTCTCAAAATTTTATTTTCAAGTGCACAAAACCAAGTTTTCCTCAAAATTTATA 5228
5345 ATCTTTGAATTAAGTATTAATCTTTTATTTTGAATTTGAGTGGAGTACCTCCATTA 5404
5229 ATCTTTGAATTAAGTATTAATCTTTTATTTTGAATTTGAGTGGAGTACCTCCATTA 5288
5405 TTAGATAAATCTTTTTCGATTAATAATCAGAAATTTTTCCTTTTCTCAAAATTTATA 5464
5289 TTAGATAAATCTTTTTCGATTAATAATCAGAAATTTTTCCTTTTCTCAAAATTTATA 5348
5465 CATATGATGATTAATATATCCATATATAGTTTTCCTGATTAATTAATGATTAATAAT 5524
5349 CATATGATGATTAATATATCCATATATAGTTTTCCTGATTAATTAATGATTAATAAT 5408
5525 AATTGGGGTCTTCAGGACCTTTTCTTATATTTTAAAGTATATCTTTTATAGCAAG 5584
5409 AATTGGGGTCTTCAGGACCTTTTCTTATATTTTAAAGTATATCTTTTATAGCAAG 5468
5585 AACATATCTGAATGTTTTTAAATCTTTTAAATTTATATGATGAGTAAATTTTGTAT 5644

5469 AACATATCTGAATGTTTTTATATAAATCTTTAAATTTATATGATAGTAAATTTTGTAT 5528
5645 CACAATGCAATTTATTTTTCCTTCTTCTTCCAACTATATACCACTGTATTTACCACTT 5704
5529 CACAATGCAATTA-TTTTTTCTTCTTCTTCCAACTATATACCACTGTATTTACCACTT 5587
5705 CTAAGAGTCACTGACACGGGCGGAGTACCCCTGAGTGTAGTCAATATGATAGCAATAAT 5764
5588 CTAAGAGTCACTGACACGGGCGGAGTACCCCTGAGTGTAGTCAATATGATAGCAATAAT 5647
5765 GAAGCCTGAAAACAGGTTTTTTTACTTTCCACTTTTAAATCCTTAGAAATTTCTTGCAACTTC 5824
5648 GAAGCCTGAAAACAGGTTTTTTTACTTTCCACTTTTAAATCCTTAGAAATTTCTTGCAACTTC 5707
5825 GCATATTTTCAATGACACCTGGTGTATAGTATAAATTTTAAATGAACTATTTACTTTTGA 5884
5708 GCATATTTTCAATGACACCGGTGTATAGTATAAATTTTAAATGAACTATTTACTTTTGA 5767
5885 TATTTTAAATCTCTTATATGTTAGTATTTTATATAACAGGATATTTAAATAGTTAAAT 5944
5768 TATTTTAAATCTCTTATATGTTAGTATTTTATATAACAGGATATTTAAATAGTTAAAT 5827
5945 CCTATGATTTGAAATTTTACAGAGCTTTCTCTTTACTTCAAAACAGCAAAAAGTGGG 6004
5828 CCTATGATTTGAAATTTTACAGAGCTTTCTCTTTACTTCAAAACAGCAAAAAGTGGG 5887
6005 GGGCATATTTGATGCTGCTGCTCATTTTAAAGTATGTAATAAATTTTAAATTTTCTGATGCT 6064
5888 GGGCATATTTGATGCTGCTGCTCATTTTAAAGTATGTAATAAATTTTAAATTTTCTGATGCT 5947
6065 TTAACATTTCTCATGTGTAATATATGTTTTTGTATCAAAAACACTCATATTTTCAAGAA 6124
5948 TTAACATTTCTCATGTGTAATATATGTTTTTGTATCAAAAACACTCATATTTTCAAGAA 6007
6125 AAAGAAATTTATGTTAAATAGCCCTGTTTTTAAAGAAATATTTTATGAAGTCTCAACTTG 6184
6008 AAAGAAATTTATGTTAAATAGCCCTGTTTTTAAAGAAATATTTTATGAAGTCTCAACTTG 6067
6185 AAGATCAAGTCAAAATTAATCACTCAGGATCTCAGGCTCTCAAGTCTAGGAGAGTCAAGAT 6244
6068 AAGATCAAGTCAAAATTAATCACTCAGGATCTCAGGCTCTCAAGTCTAGGAGAGTCAAGAT 6127
6245 TTTATCAGTTTGGGCATATAGTTTGGACTGAAATCAGATCTGTAGTCTAGTCTAGCCTAGCAAGAC 6304
6128 TTTATCAGTTTGGGCATATAGTTTGGACTGAAATCAGATCTGTAGTCTAGTCTAGCCTAGCAAGAC 6187
6305 AATTTGGAGGAGAAATATCAGCCTTCTGGAAGTGTAGTCTTCTGAAACAAATGTAAGTGTCT 6364
6188 AATTTGGAGGAGAAATATCAGCCTTCTGGAAGTGTAGTCTTCTGGAACAAATGTAAGTGTCT 6247
6365 GCAGATATTTCAATAAATGGCAACCTGTATTA 6397
6248 GCAGATATTTCAATAAATGGCAACCTGTATTA 6280

RESULT 3

HSM804398

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

HSM804398 5841 bp mRNA linear PRI 16-JUN-2003
Homo sapiens mRNA; cDNA DKFP451M1519 (from clone DKFP451M1519).
AL833087
AL833087.2 GI:30268375
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 5841)
Bloeker, H., Boecher, M., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,
Pobo, G., Han, M. and Wiemann, S.
Direct Submission
Submitted (16-JUN-2003) MIPS, Ingolstaedter Landstr.1, D-85764
Neuberberg, GERMANY
On Apr 30, 2003 this sequence version replaced gi:21733678.

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.

This clone (DKFZp451M1519) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.

FEATURES

Location/Qualifiers

1..5841

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="RZPD:DKFZp451M1519"

/db_xref="taxon:9606"

/map="5q14.1"

/clone="DKFZp451M1519"

/tissue_type="human.spinal.cord"

/clone_lib="451 (synonym: hlcc1). Vector pSport1; host DH10B; sites NotI + SalI"

/dev_stage="adult"

1..5841

/gene="DKFZp451M1519"

<1..3847

/gene="DKFZp451M1519"

/note="endofin, N-terminus truncated"

/codon_start=2

/product="hypothetical protein"

/protein_id="CAD89968.1"

/db_xref="GI:30268376"

/translation="KDKLQHSQPCGLIKDVLKKEVDVAVITAAECLKEGKTSAL
TCLPKNEDLNDNSRDENFKLPDFSFQEDKVIKQSAQEDSKSLDKNDVIOOS
SSALHVSFDEKSLSPASGSCGSLIESKARGDPLPOHEHKINQDAVITHEBIO
NSVLGEGPEKNDLQKQKSGILLOSLEGMEDKIDPDQTVIRAESLDGGDTST
VVEOGLGKHAHPDESDDCEGTTFSSNDMDQDIDYFNIDEGAKSGLIISDALSD
AFLEQYLOVQNTKISFENVDNSKQNMNMDKGLDGNINNVFNBAEAGIESHG
NLICETVDKNTLNGSLGKSTIPVQOGLPTSKSEITNQLSVINDNSOSVGAAPK
QLSLPSRTSSKDLNKPVDPTIESEPTADTVPTICALDSTADPOVFSFNVDI
ESNSESSEFVIANEDSVPTNKGVLVQKQPTWVPDSEAPNMCNKQVFTFKRR
HHCRCAGKCVGCNCKQLQVLEKARVVCVETISKAQAFERMSPTGNLKNR
HSDECTVQPOENQTSIPSPATLPVSAHQFEGVLCSEKQKRVNFDGILPNEG
ADTTKLSGSKRSEDESPSPDPVMTVNTVDHSHSTVEKPNNETGDIRDEIIQSP
ISQVPSVEKLSMTNGELPTSGSFTLLDDVDFAETEPSPPTGLVNSNLIASIDY
RLIDINKYCNKISLIPNDESLPLLVASGEKSPVVEEHPHSEHQIILLLEGSP
HPVTFVLNALLNVNKEIFYSKDYWFSTNGLEGLQQAIIILLCLNEDTIPKDI
FRLLITIKALKKYLENDNITTFESFLSSDHGFLFITPTFOKLDLSDSPNPF
LCGLIQKLEIPWAKVPMRLMLRLGAEYKAYPAPTSIEGRKPLFEIGHITMNLV
DLRNYQYTLNIDOLLHMEMGKSCIKIPRKYSVMKVLNNSNEHVISIGASFTEA
DSHLVCTQNDGIYETOANSATGHPKVTGASFVVFNGALXTSSGFLAKSIVEDGLMV
QITPTWNGRLALREOKDFKITCGKVDVLDREYVDICWVDAEENKGVISVVDGI
SLOGFPSEKILEADFTDDEKIVKCTEVFVFLKDQDLSILSTSVQFAKETAMACSAAL
CPHUKTLKSGMKNKIGURVSLDIDTWFEFQAGSEGLLPQHYLNDLSALIPVHGGTS
NSSLPLEIELVFFIIIEHLF"

ORIGIN

Query Match 84.1%; Score 5579.8; DB 9; Length 5841;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 5598; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 1022 CAAGAGCAAGCTAACAACAAGAGCCAGCCCATGTGGATTACTAAAGATGTGGCTTAGT 1081

DB 1 CAAGAGCAAGCTAACAACAAGAGCCAGCCCATGTGGATTACTAAAGATGTGGCTTAGT 60

QY 1082 AAGAGAGCAAGCTAGATGGGAGTCATCAATGCCGCGAGATGTTTAAAGAGAGGGCAA 1141

DB 61 AAGAGAGCAAGCTAGATGGGAGTCATCAATGCCGCGAGATGTTTAAAGAGAGGGCAA 120

QY 1142 GACAAGTCTTTGACCTGCAGCCTTCGAAAAATGAAGATTTATGCTTTAAATGATTCAAA 1201

DB 121 GACAAGTCTTTGACCTGCAGCCTTCGAAAAATGAAGATTTATGCTTTAAATGATTCAAA 180

QY 1202 TTCAAGAGATGAAATTTCAATTTACCTGACTTTTCCTTTTCAGGAAGATAAGACTGTAT 1261

181 TTCAAGAGATGAAATTTCAATTTACCTGACTTTTCCTTTTCAGGAAGATAAGACTGTAT 240

1262 AAAAAAATCTGCACAGAGAGACTCAAAAAGTTTACAGCTTACAGGTAATGATGTAATCCA 1321

241 AAAAAAATCTGCACAGAGAGACTCAAAAAGTTTACAGCTTACAGGTAATGATGTAATCCA 300

1322 AGATTCCTCTTTCAGCTTTTACATGTTTCCAGTAAAGATGTGCGCTCCTCATTTCTCTGTCT 1381

301 AGATTCCTCTTTCAGCTTTTACATGTTTCCAGTAAAGATGTGCGCTCCTCATTTCTCTGTCT 360

1382 TCTCGGCTCGGCTCTATGTGTGGATCATTATTGAAGTAAAGACACGGGGTGAATTTT 1441

361 TCTCGGCTCGGCTCTATGTGTGGATCATTATTGAAGTAAAGACACGGGGTGAATTTT 420

1442 ACCTCAGCATGAACATAAAGATAATATACAAGATGACAGTATACATGAAAGAAATACA 1501

421 ACCTCAGCATGAACATAAAGATAATATACAAGATGACAGTATACATGAAAGAAATACA 480

1502 GAACAGTGTGTTTACAGTGGGGAACCATTAAGAGAGAAATGATCTTTTGAAGACAGAAAA 1561

481 GAACAGTGTGTTTACAGTGGGGAACCATTAAGAGAGAAATGATCTTTTGAAGACAGAAAA 540

1562 ATGTAAGCATACCTCTTCAGTCAATTAATGAAGGATGGAAGACAGAAAGATAGATCC 1621

541 ATGTAAGCATACCTCTTCAGTCAATTAATGAAGGATGGAAGACAGAAAGATAGATCC 600

1622 TGACCAAGCAGTAATCAGAGCTGAGTCTTTTGGATGGTGTGACACAGTCTTACAGTTGT 1681

601 TGACCAAGCAGTAATCAGAGCTGAGTCTTTTGGATGGTGTGACACAGTCTTACAGTTGT 660

1682 AGAATCTCAAGAGGGCTTTCTGGCAGTCACTCTCCAGAGTCTTCTGATGTTGTGAAGG 1741

661 AGAATCTCAAGAGGGCTTTCTGGCAGTCACTCTCCAGAGTCTTCTGATGTTGTGAAGG 720

1742 TTTTATTAATACTTTTCAAGCAATGATATGATGGGCAAGACTTAGATTACTTTAATAT 1801

721 TTTTATTAATACTTTTCAAGCAATGATATGATGGGCAAGACTTAGATTACTTTAATAT 780

1802 TGATGAAGCGCAAAAAGTGGCCCACTAATTAAGTGTCTGAGTCTTGTGCTTCTTGAC 1861

781 TGATGAAGCGCAAAAAGTGGCCCACTAATTAAGTGTCTGAGTCTTGTGCTTCTTGAC 840

1862 AGAAGCAGTATCTTCAGACCACTTAAACATAAAGTCTTTTGAAGAAAATGTAATGACTCTAA 1921

841 AGAAGCAGTATCTTCAGACCACTTAAACATAAAGTCTTTTGAAGAAAATGTAATGACTCTAA 900

1922 ATCGCAATGAATCAGATAGATATGAAGGCTTAGATGATGGAACATCAATAATATATA 1981

901 ATCGCAATGAATCAGATAGATATGAAGGCTTAGATGATGGAACATCAATAATATATA 960

1982 TTTCAATGCAAGCAGGAGGACTATTGGGGAAGTCAATGATTAATATAATTTGTGAAC 2041

961 TTTCAATGCAAGCAGGAGGACTATTGGGGAAGTCAATGATTAATATAATTTGTGAAC 1020

2042 AGTTGATAAACAATAACATAAGAAATGGCCCTTTCTTTAGGAGAAAAAGCACTATTCC 2101

1021 AGTTGATAAACAATAACATAAGAAATGGCCCTTTCTTTAGGAGAAAAAGCACTATTCC 1080

2102 AGTTCAACAGGGTACCTTACCAAGTGTCTGAGATTACAAATCAATTCAGTCTCTCA 2161

1081 AGTTCAACAGGGTACCTTACCAAGTGTCTGAGATTACAAATCAATTCAGTCTCTCA 1140

2162 TATTACAGTCAATCTGTTGGAGGGGCCAGACCTTAAGCAATTTGTTAGCCTTCCATCAAG 2221

1141 TATTACAGTCAATCTGTTGGAGGGGCCAGACCTTAAGCAATTTGTTAGCCTTCCATCAAG 1200

2222 ACAAGAGGTTCAAAGGACCTGAATAAGCCAGATGTTCCAGATACAAATGAAGTGAACC 2281

1201 ACAAGAGGTTCAAAGGACCTGAATAAGCCAGATGTTCCAGATACAAATGAAGTGAACC 1260

2282 CAGCAGCAGATACCGTTGTTTCCCAATCACTTGTGTATAGATTCTACAGCTGATCCACA 2341

Db 1261 CAGCAGCAGATACCGTGTGTTCCATCAGTGTGCTATAGATTTCTACAGCTGATCCACA 1320
Qy 2342 GGTAGCTTCAACTCTAATACATGATPATAGAAAGTAATTTCTGAAGGTGATCTAGTTT 2401
Db 1321 GGTAGCTTCAACTCTAATACATGATPATAGAAAGTAATTTCTGAAGGTGATCTAGTTT 1380
Qy 2402 CGTACTCAATGAGATTCGTACCTGGAACACACTTGCAGAAAGAGCTGTTGTTGGG 2461
Db 1381 CGTAACTCAATGAGATTCGTACCTGGAACACACTTGCAGAAAGAGCTGTTGTTGGG 1440
Qy 2462 CCAGAAACAGCTACTTGGGTTCCCTGATTGAGAGCTCCAACTGTTGAACTGCCAAGT 2521
Db 1441 CCAGAAACAGCTACTTGGGTTCCCTGATTGAGAGCTCCAACTGTTGAACTGCCAAGT 1500
Qy 2522 CAAATTTACTTTTACCAACCGGAGCAGCATGCGGAGCATGTGGGAAAGTATTTTGGG 2581
Db 1501 CAAATTTACTTTTACCAACCGGAGCAGCATGCGGAGCATGTGGGAAAGTATTTTGGG 1560
Qy 2582 TGTCTGTTGTAATAGGAAGTGTAACTGCAATATCTAGAAAGGAAGCAAGAGTATGTGT 2641
Db 1561 TGTCTGTTGTAATAGGAAGTGTAACTGCAATATCTAGAAAGGAAGCAAGAGTATGTGT 1620
Qy 2642 AGTCTGCTATGAACCTATTAGTAAAGCTCAGGCAATTTGAAAGGATGATAGTCCAACTGG 2701
Db 1621 AGTCTGCTATGAACCTATTAGTAAAGCTCAGGCAATTTGAAAGGATGATAGTCCAACTGG 1680
Qy 2702 TTCTAATCTTAAGTCTAATCATCTGATGAATGTACTACTGTCAGGCTCTCTCAGGAGAA 2761
Db 1681 TTCTAATCTTAAGTCTAATCATCTGATGAATGTACTACTGTCAGGCTCTCTCAGGAGAA 1740
Qy 2762 CCAACATCCAGTATACCTTCCACAGCAACTTCCAGTCTCAGCACTTAAACAAACCAGG 2821
Db 1741 CCAACATCCAGTATACCTTCCACAGCAACTTCCAGTCTCAGCACTTAAACAAACCAGG 1800
Qy 2822 TGTGAAGGACTATGTTTCCAAAGACAGAGAGATGTTTCCAGATGATATTTGCC 2881
Db 1801 TGTGAAGGACTATGTTTCCAAAGACAGAGAGATGTTTCCAGATGATATTTGCC 1860
Qy 2882 CAAATGCTGAAGTGCAGATACAAACAAATTTATCTTGGAAAGTAAAGATGTTCTGAAGA 2941
Db 1861 CAAATGCTGAAGTGCAGATACAAACAAATTTATCTTGGAAAGTAAAGATGTTCTGAAGA 1920
Qy 2942 CTTTACTCTCTCCTCAGCTGATGCTGATGACAGTAAACACAGTGAATCTCCCAATTC 3001
Db 1921 CTTTACTCTCTCCTCAGCTGATGCTGATGACAGTAAACACAGTGAATCTCCCAATTC 1980
Qy 3002 TACTACAGTGGAAAGCCAAACATGACAGAGGAGATATTAACAGAAATGAGATAATTCA 3061
Db 1981 TACTACAGTGGAAAGCCAAACATGACAGAGGAGATATTAACAGAGATGAGATAATTCA 2040
Qy 3062 GAGTCTTATTTCTCAGGTTCCATCAGTGGAAATTTGTTGATGAACACAGGAATGAGGG 3121
Db 2041 GAGTCTTATTTCTCAGGTTCCATCAGTGGAAATTTGTTGATGAACACAGGAATGAGGG 2100
Qy 3122 GTTACTCTACTCTGTTTCATTTACATGATGATGATGTTTTCGAGAACTGAGGAACC 3181
Db 2101 GTTACTCTACTCTGTTTCATTTACATGATGATGATGTTTTCGAGAACTGAGGAACC 2160
Qy 3182 ATCTAGTCTCTACTGTTGTTCTAGTTTAAACAGCAATTTTACCTATTGCTAGTATTTCAAGATTA 3241
Db 2161 ATCTAGTCTCTACTGTTGTTCTAGTTTAAACAGCAATTTTACCTATTGCTAGTATTTCAAGATTA 2220
Qy 3242 TAGGTTACTGTGATATTAACAGTATGCTGCAATAGATTTAGTCTTCTACCTAATGA 3301
Db 2221 TAGGTTACTGTGATATTAACAGTATGCTGCAATAGATTTAGTCTTCTACCTAATGA 2280
Qy 3302 TGAGGACAGTTTGGCCCCACATTTCTGGTTCATCTGGAAGAAAGGATCAGTGCCTGTAGT 3361
Db 2281 TGAGGACAGTTTGGCCCCACATTTCTGGTTCATCTGGAAGAAAGGATCAGTGCCTGTAGT 2340
Qy 3362 AGAAGAACATCTCATGAGCAGATCATTTTGGTTCCTGAAAGGTGAAGGCTTTTATCC 3421
Db 2341 AGAAGAACATCTCATGAGCAGATCATTTTGGTTCCTGAAAGGTGAAGGCTTTTATCC 2400

Qy 3422 TGTTACATTTTGTCTTAATGCTTAATCTACTCGTGAATGTCAAATTCATATTTTATTCCTC 3481
Db 2401 TGTTACATTTTGTCTTAATGCTTAATCTACTCGTGAATGTCAAATTCATATTTTATTCCTC 2460
Qy 3482 AGACAAATATTTGGTACTTTTCAACCAATGGATTGCTTTGGGACAGGACAGAAATTTAT 3541
Db 2461 AGACAAATATTTGGTACTTTTCAACCAATGGATTGCTTTGGGACAGGACAGAAATTTAT 2520
Qy 3542 TATTTCTATTTGTTATGTTTGGCCAAATGAAGATACTATTCCTAAGGACATCTTCAGACTATT 3601
Db 2521 TATTTCTATTTGTTATGTTTGGCCAAATGAAGATACTATTCCTAAGGACATCTTCAGACTATT 2580
Qy 3602 TATCACCATATATAGGATGCTCTAAAGGAAATACATAGAAAACCTTGGACAAATATTAC 3661
Db 2581 TATCACCATATATAGGATGCTCTAAAGGAAATACATAGAAAACCTTGGACAAATATTAC 2640
Qy 3662 CTTTACTGAGAGTTTCTCAGTAGCAAGATCACGAGAGATTCCTGTTTATTTACACCTAC 3721
Db 2641 CTTTACTGAGAGTTTCTCAGTAGCAAGATCACGAGAGATTCCTGTTTATTTACACCTAC 2700
Qy 3722 TTTTCAGAACTTGATGATCTCTCATTTACCAGTAACTCTTTCTTTCTTTGGAATTTCTTAT 3781
Db 2701 TTTTCAGAACTTGATGATCTCTCATTTACCAGTAACTCTTTCTTTCTTTGGAATTTCTTAT 2760
Qy 3782 CCAGAGAGCTTGAGATTTCCCTGGGCAAGGTTTTTCTATGCGTTTTTAATTTGAGATTGGG 3841
Db 2761 CCAGAGAGCTTGAGATTTCCCTGGGCAAGGTTTTTCTATGCGTTTTTAATTTGAGATTGGG 2820
Qy 3842 TGCAGAAATTAAGCATATCTCTCTCTTAACAGCATCAGAGCCGGAACCTTTT 3901
Db 2821 TGCAGAAATTAAGCATATCTCTCTCTCTTAACAGCATCAGAGCCGGAACCTTTT 2880
Qy 3902 TGCAGAAATTAAGCATATCTCTCTCTCTTAACAGCATCAGAGCCGGAACCTTTT 3961
Db 2881 TGCAGAAATTAAGCATATCTCTCTCTCTTAACAGCATCAGAGCCGGAACCTTTT 2940
Qy 3962 CTTGCAATATATAGATCAACTCTGTTGATGCAATGGAATGGAAGAAAGCTGCAATAAAAAT 4021
Db 2941 CTTGCAATATATAGATCAACTCTGTTGATGCAATGGAATGGAAGAAAGCTGCAATAAAAAT 3000
Qy 4022 ACCACGGAAGAAAGTACAGTATGTAATGAAGTACTTAAATTTCTTCAATGAGCATGTCTAT 4081
Db 3001 ACCACGGAAGAAAGTACAGTATGTAATGAAGTACTTAAATTTCTTCAATGAGCATGTCTAT 3060
Qy 4082 TAGCATTTGGAGCAAGTTTTCAGTACAGAGCAGATTTCTCATCTAGTCTGTATACAGAAATGA 4141
Db 3061 TAGCATTTGGAGCAAGTTTTCAGTACAGAGCAGATTTCTCATCTAGTCTGTATACAGAAATGA 3120
Qy 4142 TGGAAATTTATGAACACAGGCCAACAGTCCCACTGCGCATCTTCAAGAAAGTACAGAGTGC 4201
Db 3121 TGGAAATTTATGAACACAGGCCAACAGTCCCACTGCGCATCTTCAAGAAAGTACAGAGTGC 3180
Qy 4202 AAGTTTTGTGGTATTTCAATGAGCTCTTAAACACATCTTTCAGGATTTCTTGTAGTCCAG 4261
Db 3181 AAGTTTTGTGGTATTTCAATGAGCTCTTAAACACATCTTTCAGGATTTCTTGTAGTCCAG 3240
Qy 4262 CATAGTTGAAGATGCTTTAATGTTACAAATACTCCAGAGACCATGAAATGCTTCCGCT 4321
Db 3241 CATAGTTGAAGATGCTTTAATGTTACAAATACTCCAGAGACCATGAAATGCTTCCGCT 3300
Qy 4322 AGCTTTTCCAGAACAGAAAGACTTTTAAATTTACATGTTGGAAAGTTGATGCAAGTACAGCT 4381
Db 3301 AGCTTTTCCAGAACAGAAAGACTTTTAAATTTACATGTTGGAAAGTTGATGCAAGTACAGCT 3360
Qy 4382 GAGAGAAATCGTGGATATCTGCTGGTGTAGTCTGAGAAAGAAAGCAAGAGGATTTAT 4441
Db 3361 GAGAGAAATCGTGGATATCTGCTGGTGTAGTCTGAGAAAGAAAGCAAGAGGATTTAT 3420
Qy 4442 CAGTTTCAGTGGATGGAATATCATTTACAGGATTTTCCAAAGTAAAAAATAAACTTGGAAAGC 4501
Db 3421 CAGTTTCAGTGGATGGAATATCATTTACAGGATTTTCCAAAGTAAAAAATAAACTTGGAAAGC 3480

QY 4502 AGATTTTGAACCGATGAGAGATTTGTAATGTCACGGGTGTTCTACTTCTTAAGGA 4561
Db |||||
QY 3481 AGATTTTGAACCGATGAGAGATTTGTAATGTCACGGGTGTTCTACTTCTTAAGGA 3540
Db |||||
QY 4562 CCAGGATTTATCTATTTTATCAAACTTCTTATCAGTTTGCAAAAGAAATAGCCATGCGCTTG 4621
Db |||||
QY 3541 CCAGGATTTATCTATTTTATCAAACTTCTTATCAGTTTGCAAAAGAAATAGCCATGCGCTTG 3600
QY 4622 TAGTGTGCGCTGTGCGCTCACCTGAAACCTCTAAAGTAATGGGATGAATTAATTTGG 4681
Db |||||
QY 3601 TAGTGTGCGCTGTGCGCTCACCTGAAACCTCTAAAGTAATGGGATGAATTAATTTGG 3660
QY 4682 ACTCAGAGTTTCCATTCACACTGATATGGTTGTAATTCAGGCAGGATCTGAAGGCCAACT 4741
Db |||||
QY 3661 ACTCAGAGTTTCCATTCAGACTGATATGGTTGTAATTCAGGCAGGATCTGAAGGCCAACT 3720
QY 4742 TCTGCTCAGCAATATCTAAATGATCTTGATAGTGTCTCATACCTGTGATCCATGGTGG 4801
Db |||||
QY 3721 TCTGCTCAGCAATATCTAAATGATCTTGATAGTGTCTCATACCTGTGATCCATGGTGG 3780
QY 4802 GACCTCCAACTCTAGTTTACCAATAGAAATAGAAATAGTGTCTTTCATATAGCAATCT 4861
Db |||||
QY 3781 GACCTCCAACTCTAGTTTACCAATAGAAATAGAAATAGTGTCTTTCATATAGCAATCT 3840
QY 4862 TTTTGTAGTGAAGAAATGTGCCATATTAATATGCAACCTTAATTTGTTAAACCTAACTCC 4921
Db |||||
QY 3841 TTTTGTAGTGAAGAAATGTGCCATATTAATATGCAACCTTAATTTGTTAAACCTAACTCC 3900
QY 4922 AGCACTAAAGCTGAATGCGCAACAACTAAAGATTAATATGCTGATTTTGAACA 4981
Db |||||
QY 3901 AGCACTAAAGCTGAATGCGCAACAACTAAAGATTAATATGCTGATTTTGAACA 3960
QY 4982 CATAGCTTTGCTCTTTAGGAGGAATGATCTTTCAAATCAATAGCACAAATTTAAAT 5041
Db |||||
QY 3961 CATAGCTTTGCTCTTTAGGAGGAATGATCTTTCAAATCAATAGCACAAATTTAAAT 4020
QY 5042 ATCTAAATTTAAGAGATCCTACTTCTGTAGCTTTCAATTAATTTAAGTACTAAAT 5101
Db |||||
QY 4021 ATCTAAATTTAAGAGATCCTACTTCTGTAGCTTTCAATTAATTTAAGTACTAAAT 4080
QY 5102 AGCAAGGATTTCTTTTGAAGAAATTAATAGCAATTTACTGTGTTATTTTAAATGCTAAGCCA 5161
Db |||||
QY 4081 AGCAAGGATTTCTTTTGAAGAAATTAATAGCAATTTACTGTGTTATTTAATGCTAAGCCA 4140
QY 5162 AAGTATCTGCACCTTAGGTATACCTCTTTATGCCAATTAATGATTTTAAATGAAGCTCTTTT 5221
Db |||||
QY 4141 AAGTATCTGCACCTTAGGTATACCTCTTTATGCCAATTAATGATTTTAAATGAAGCTCTTTT 4200
QY 5222 CAGATGTAACTTATGAAGGAATATCTGCTTTGTGTATATGCGAGTTAGAACTAGTGT 5281
Db |||||
QY 4201 CAGATGTAACTTATGAAGGAATATCTGCTTTGTGTATATGCGAGTTAGAACTAGTGT 4260
QY 5282 TCTAAAGTCTGTCAAATGTTATTTTCACTGGGCACAAAACCAAGTTTGGAGTCTTAGACTT 5341
Db |||||
QY 4261 TCTAAAGTCTGTCAAATGTTATTTCACTGGGCACAAAACCAAGTTTGGAGTCTTAGACTT 4320
QY 5342 ATAAATCTTTGAATAAACCTGATAACTTTATGTAATTTGGAGTGGAGACCTACCTCCA 5401
Db |||||
QY 4321 ATAAATCTTTGAATAAACCTGATAACTTTATGTAATTTGGAGTGGAGACCTACCTCCA 4380
QY 5402 TAAATAGATAAATCTTTTGGATTAATTAATCAGAAATTTTGCCTTTTCTCTCAAATTA 5461
Db |||||
QY 4381 TAAATAGATAAATCTTTTGGATTAATTAATCAGAAATTTTGCCTTTTCTCTCAAATTA 4440
QY 5462 TTACATATGATGATTTATATATATCCATATATAGTTTCCCTGATTAATGATATTA 5521
Db |||||
QY 4441 TTACATATGATGATTTATATATATCCATATATAGTTTCCCTGATTAATGATATTA 4500
QY 5522 AATAATTTGCGGCTCTCAGACCTTTTGTCTCTATATTTAAGTATATCTTTTATAGC 5581
Db |||||
QY 4501 AATAATTTGCGGCTCTCAGACCTTTTGTCTCTATATTTAAGTATATCTTTTATAGC 4560
QY 5582 AAGAACAATTTCTGAATGTTTATAAATCTTTTAAATTTATATGATAGTAAATTTTGG 5641

RESULT 4

4561 AAGAACAATTTCTGAATGTTTATAAATCTTTAATATTTATATGATAGTAAATTTTGG 4620
QY 5642 TATCAAAATGCAATTAATTTTCTCTCTTCTTCCAACTATACCACTGATTTTACCA 5701
Db |||||
QY 4621 TATCAAAATGCAATTAATTTTCTCTCTTCTTCCAACTATACCACTGATTTTACCA 4679
QY 5702 CTTCTTAAGAGTGAATGACGAGCGGCAGATGACCTTGAAGTAGTCAATTTATGTAACAATA 5761
Db |||||
QY 4680 CTTCTTAAGAGTGAATGACGAGCGGCAGATGACCTTGAAGTAGTCAATTTATGTAACAATA 4739
QY 5762 AATGAAGCTGAAACAGAGTTTCTTCTCACTTTAATCTTGAATAATTTCTTGGCAAC 5821
Db |||||
QY 4740 AATGAAGCTGAAACAGAGTTTCTTCTCACTTTAATCTTGAATAATTTCTTGGCAAC 4799
QY 5822 TCTGCAATTTTCAATGACACTGGGTATAGTATAAATTTAATGAATTAATTTACTTTT 5881
Db |||||
QY 4800 TCTGCAATTTTCAATGACACTGGGTATAGTATAAATTTAATGAATTAATTTACTTTT 4859
QY 5882 GCATATTTTAAATCTTTTATATGTTAGTTATTTTATAACAGGATATTAACATAAGTTA 5941
Db |||||
QY 4860 GCATATTTTAAATCTTTTATATGTTAGTTATTTTATAACAGGATATTAACATAAGTTA 4919
QY 5942 AATCTATGTTTGAATTTGTTACAGAGCTTCTCTTCTTCAACAGCAAAAAGT 6001
Db |||||
QY 4920 AATCTATGTTTGAATTTGTTACAGAGCTTCTCTTCTTCAACAGCAAAAAGT 4979
QY 6002 GGGGGCATATTTGTTAGTCTCTGTCATTTAAGTTATCTTAAATAATTTAATCAATTTTGT 6061
Db |||||
QY 4980 GGGGGCATATTTGTTAGTCTCTGTCATTTAAGTTATGTAATAAATCTAATCAATTTTGT 5039
QY 6062 GCTTTTAAACATTTCTCATGTGTAATATATGTTTGTATCAAAAACACTCATATATTTCAA 6121
Db |||||
QY 5040 GCTTTTAAACATTTCTCATGTGTAATATATGTTTGTATCAAAAACACTCATATATTTCAA 5099
QY 6122 GAABAAGAAATTTATGTTTAAATAGCCCTGTTTAAAGAAATATTTATGAAGCATCTCAAC 6181
Db |||||
QY 5100 GAABAAGAAATTTATGTTTAAATAGCCCTGTTTAAAGAAATATTTATGAAGCATCTCAAC 5159
QY 6182 TTGAAGATCAAGTCAAAAGTTATAACTCAGGATCTCAGGTCTCAAGCTAGGAGAGACTGAG 6241
Db |||||
QY 5160 TTGAAGATCAAGTCAAAAGTTATAACTCAGGATCTCAGGTCTCAAGCTAGGAGAGACTGAG 5219
QY 6242 AATTTTAAATCAGTTTGGGCATATAGTTTGGAGTGAATCACATCTCTAGTACTTACCCAAA 6301
Db |||||
QY 5220 AATTTTAAATCAGTTTGGGCATATAGTTTGGAGTGAATCACATCTCTAGTACTTACCCAAA 5279
QY 6302 GACAAATTTGGAGGAGATATCAGCCCTTCTGGAAGTAGTCTTCTCTGAACATGTAAGT 6361
Db |||||
QY 5280 GACAAATTTGGAGGAGATATCAGCCCTTCTGGAAGTAGTCTTCTCTGAACATGTAAGT 5339
QY 6362 GTGCGAGATTTCAATAAATGGCAACCTGTTTATTAATTTGTGAATTTTATGAAATGGTG 6421
Db |||||
QY 5340 GTGCGAGATTTCAATAAATGGCAACCTGTTTATTAATTTGTGAATTTTATGAAATGGTG 5399
QY 6422 TAAGATGAACCAATTTGCAATCAAAACCAATTTTATTTCTTAAATAGTGTATGAT 6481
Db |||||
QY 5400 TAAGATGAACCAATTTGCAATCAAAACCAATTTTATTTCTTAAATAGTGTATGAT 5459
QY 6482 TCTGCCATGTAAGTAAATGAACAGTCTTAAAAATAACCAATGGTAGAGGCTGTTCCATG 6541
Db |||||
QY 5460 TCTGCCATGTAAGTAAATGAACAGTCTTAAAAATAACCAATGGTAGAGGCTGTTCCATG 5519
QY 6542 ATGGGACAGCTTTGGATTTGTTTCTAATAAATCTCTACATCAATAAATAATGGAATTAAT 6601
Db |||||
QY 5520 ATGGGACAGCTTTGGATTTGTTTCTAATAAATCTCTACATCAATAAATAATGGAATTAAT 5579
QY 6602 GTGCGCTGAAGTTTGGAGGCACATTTTGAAGT 6632
Db |||||
QY 5580 GTGCGCTGAAGTTTGGAGGCACATTTTGAAGT 5610

AF434817 LOCUS 4620 bp mRNA linear PRI 12-NOV-2001
 DEFINITION Homo sapiens endofin mRNA, complete cds.
 ACCESSION AF434817
 VERSION AF434817.1 GI:16904135
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 4620)
 SEED L.F. and Hong, W.
 TITLE Endofin, an endosomal FYVE domain protein
 JOURNAL J. Biol. Chem. (2002) In press
 REFERENCE 2 (bases 1 to 4620)
 AUTHORS Seet, L.F. and Hong, W.
 TITLE Direct Submission
 JOURNAL Submitted (17-OCT-2001) Membrane Biology Laboratory, Institute of
 Mol & Cell Biol, 30 Medical Drive, Singapore 117609, Singapore
 FEATURES
 Location/Qualifiers
 1..4620
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 1..4620
 /note="endosomal FYVE domain protein"
 /codon_start=1
 /product="endofin"
 /protein_id="AAL30772.1"
 /db_xref="GI:16904135"
 /translation="MDSYFKAASVDLKLDDPFDPEODYLDQVQNAVDSNHCVS
 SELASSQRTSLPKDQSCVNSCASSETSYGTSNLSNEKTLKGLTSTQNEKNVTGLDL
 LSSVDGSDIQPLVWGRSKPICDLISMGNLVHTATNEEDIKLLPDPFKSNADS
 LIGLDSVSDTDDSDHSDRLQSQNDTSELQNEIGGKELGKIVDTITLSDS
 YNYSGETNVAITAECLKEGKGTALTCPQNEDELCLNDSNEDENFKLPDFSFED
 LVKEEDVAVTAECLKEGKGTALTCPQNEDELCLNDSNEDENFKLPDFSFED
 KTVIKQAQEDKSLDKNDVIQDSSALHVS KDVPSSLSCLPASGCMGSLIESK
 ARGDFLQAEHDKNIQDAVTETHEIIONSVLVGGPKFNDLLKQEKCSLLQSLLEG
 MEDKIPDPQVPIAESLGDGDSSTVSESQGLSGTHVPESDCEGFINTESSNDM
 DQQLDFINDEGAKSGLSDAELDAFLTEYQLQTNIKSPFENVYDSSKSNQIDM
 KGLDGGNNININFAEAGAGESGINIICETVDKQNTIENGSLSEKSTIPVQQLP
 TSSEITINQSVSDINSQSGARPQLFSLPGRSTSSDLNKPDPDTIESPESFAD
 TVPITTCALDADPOVFSNYSYDIENSSEGSFVTANEDSVPTNCKEGLVLCQ
 QTVVPDSEAPNCMCOVKFTFKRHRCACGVFCVCCNCKQLQYLEKEARVCV
 VCYETISKAQAFERMSPTGSLNLSNEDCTVPOPENOTSSIIPSLPLVSAKQ
 PGVGLCKSEKRYWFADGILPNGEVADTTKLSGSKRCSEDPSPSPVPMVTNVD
 HSHSTVEKPNNEVDITREILIQSPISQVPSVEKLSMNTGNEGLPTSGSFLLDDVF
 AETEPSPTGLVNLPIASISDYELLCIDINKYCNKISSLPNEDSDSLPPLVASG
 EKGSPVVEEHPHEQIILLLEGEPHPTVFLVNLNLLVNVKFIYSSDKYWFSTNG
 LHGLQAEIILLCLNEDTIPKDIPLFLITKDALNGKYLENDNITFTFESFSS
 KDHGFLPTFTKQDLDLSPNPLFCGLIKLEIPWAKVPMRLMLRLGAEYKAY
 PAPLTSIRGRPLFGEIGHTIMNLLVLRNVQVTLNIDOLLHMEMGSKCIKPKK
 YSDVMKVLNSNEHVI SIGASFTEADSHLVCLONGIYETQANSATGHPKVTGASF
 VYFNGALKTSGLFLAKSIVSVDGLMWQITPETHNGLRLAIREQKFIITCGKVDVLD
 REYVDICWDVAENKNGVSISSVDGTSIQGFPPSEKILEADFTDSEKIVKTEVFYFL
 KDQDLISLTSYQPAKINWASALCPHLKLSNGMKNIGLRVSDITDMWFQFAS
 EGQLLPQHYLNDLSALIPVHGTSNLSPLFELVFFIIEHLF"

ORIGIN

Query Match 69.6%; Score 4615.2; DB 9; Length 4620;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 4617; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 249 ATGACAGTATTTAAAGCAGCTGTGACGACTTGGACAACTCTTGTGATTTTGA 308
 1 ATGACAGTATTTAAAGCAGCTGTGACGACTTGGACAACTCTTGTGATTTGAA 60
 309 CAGAACCCAGATGAACAGATTATCTCGCAGATGTACAAATGCATATGATTTCAACC 368
 61 CAGAACCCAGATGAACAGATTATCTCGCAGATGTACAAATGCATATGATTTCAACC 120
 369 TGCTCAGTTTCTTCCAGAGTTGGCTTCTCTCAGCGCAACTCAITGCTCCCAAGACCAA 428

121 TGCTCAGTTTCTTCCAGAGTTGGCTTCTCTCAGCGCAACTCTTCTCTCCCAAGACCAA 180
 429 GAGTGGTGTAAATAGTTGTGGCTCTCATCAGAAACAAGCTATGGAACAAATGAGAGTTCCCTG 488
 181 GAGTGGTGTAAATAGTTGTGGCTCTCATCAGAAACAAGCTATGGAACAAATGAGAGTTCCCTG 240
 489 AATGAAAAAACHACTCAAGGACTTACTTCTATACAAAATGAAAAATGTAACAGGACTT 548
 241 AATGAAAAAACHACTCAAGGACTTACTTCTATACAAAATGAAAAATGTAACAGGACTT 300
 549 GATCTTCTTCTTCTGTGTGATGGTGGTACTTCCAGATGAAATCCAGCCGCTTATATATGGA 608
 301 GATCTTCTTCTTCTGTGTGATGGTGGTACTTCCAGATGAAATCCAGCCGCTTATATATGGA 360
 609 CGATGTAGTAAACCTATCTGTGATCTGATAGTACATGGGTAACTTGTAGTTCATGCAACC 668
 361 CGATGTAGTAAACCTATCTGTGATCTGATAGTACATGGGTAACTTGTAGTTCATGCAACC 420
 669 AATAGTGAAGAGATATTAAAAAATTATTCAGATGATTTTAACTCTAAATGCAGATTCC 728
 421 AATAGTGAAGAGATATTAAAAAATTATTCAGATGATTTTAACTCTAAATGCAGATTCC 480
 729 TTGATTGGATTGGATTATCTTTCAGTGTGATCTCCCTGTGTGTTTCTTCAACAGACCAT 788
 481 TTGATTGGATTGGATTATCTTTCAGTGTGATCTCCCTGTGTGTTTCTTCAACAGACCAT 540
 789 GATAGTGATCTGTGAGAAACAAGATGATATCAGTCTCTGATTAACAATAGAGAA 848
 541 GATAGTGATCTGTGAGAAACAAGATGATATCAGTCTCTGATTAACAATAGAGAA 600
 849 ATCGGAGGAATCAAGAATTTGCGTATATAAAGTAGATACAAACACTTTCAGATTCTTATAAT 908
 601 ATCGGAGGAATCAAGAATTTGCGTATATAAAGTAGATACAAACACTTTCAGATTCTTATAAT 660
 909 TACAGTGGAAACAGAAAAATTTAAAGATAAAGATCTTTTATCAGTTAGATCAATTTGTT 968
 661 TACAGTGGAAACAGAAAAATTTAAAGATAAAGATCTTTTATCAGTTAGATCAATTTGTT 720
 969 GATTTTAACTGTCTATCTGCTTGTGATCGACAAAGTTCGAAATGTTTTCATGCCAAGAC 1028
 721 GATTTTAACTGTCTATCTGCTTGTGATCGACAAAGTTCGAAATGTTTTCATGCCAAGAC 780
 1029 AAGCTACAAACAAGACGACGACCATGTGATTAATAAAGATGTTGGCTTAGTAAAGAG 1088
 781 AAGCTACAAACAAGACGACGACCATGTGATTAATAAAGATGTTGGCTTAGTAAAGAG 840
 1089 GAAGTAGATGTGGCAGTCTAATCTGCCGACAGATGTTTAAAGAGAGGCGCAAGCAAGT 1148
 841 GAAGTAGATGTGGCAGTCTAATCTGCCGACAGATGTTTAAAGAGAGGCGCAAGCAAGT 900
 1149 GCTTTGACCTCGAGCGCTTCCGAAAAATGAAGATTTATCTCTTAAATGATTCAAAATCAAGA 1208
 901 GCTTTGACCTCGAGCGCTTCCGAAAAATGAAGATTTATCTCTTAAATGATTCAAAATCAAGA 960
 1209 GATGAAAAATTTCAATTAACCTGACATTTTCCCTTTCAGGAGATGACGTTTATTAACAA 1268
 961 GATGAAAAATTTCAATTAACCTGACATTTTCCCTTTCAGGAGATGACGTTTATTAACAA 1020
 1269 TCTGCACAAGAGACTCAAAAAGTTTACACCTTTAGGATAATGATGTAATCCAAAGATTCC 1328
 1021 TCTGCACAAGAGACTCAAAAAGTTTACACCTTTAGGATAATGATGTAATCCAAAGATTCC 1080
 1329 TCTCAGCTTTACATGTTTCCAGTAAAGATGTGCGGTCTCTCATTTGCTCTGCTTCTCTCGG 1388
 1081 TCTCAGCTTTACATGTTTCCAGTAAAGATGTGCGGTCTCTCATTTGCTCTGCTTCTCTCGG 1140
 1389 TCTGGGTCTATGTGTGGATCAATTAATGAAAAGTAAAGCAGCGGGTGAATTTTACCTCAG 1448
 1141 TCTGGGTCTATGTGTGGATCAATTAATGAAAAGTAAAGCAGCGGGTGAATTTTACCTCAG 1200
 1449 CATGAACATAAAGATAATATCAAGATGCGATGATATACATGAAGAATAACAGAACAGT 1508
 1201 CATGAACATAAAGATAATATCAAGATGCGATGATATACATGAAGAATAACAGAACAGT 1260

QY 1509 GTTGTCTAGTGGGAAACCAATCAAGAGAAATGATCTTTTGAACAGGAAATATAA 1568
Db 1261 GTTGTCTAGTGGGAAACCAATCAAGAGAAATGATCTTTTGAACAGGAAATATAA 1320
QY 1569 AGCATCTCTTCAGTCAATTAATGAGGATGGAAGAGATAGATCTCTGACCAG 1628
Db 1321 AGCATCTCTTCAGTCAATTAATGAGGATGGAAGAGATAGATCTCTGACCAG 1380
QY 1629 ACAGTAATCAGAGCTAGTCTTTGGATGGTGGTGCACACCAAGTCTCAGTGTGAGAACT 1688
Db 1381 ACAGTAATCAGAGCTAGTCTTTGGATGGTGGTGCACACCAAGTCTCAGTGTGAGAACT 1440
QY 1689 CAAGAGGGCTTTCTGCACTCATGTCCTGCAAGTCTCTGATTTGTTGAGAGTTTAT 1748
Db 1441 CAAGAGGGCTTTCTGCACTCATGTCCTGCAAGTCTCTGATTTGTTGAGAGTTTAT 1500
QY 1749 AATACTTTTTCAAGCAATGATATGGATGGGCAAGACTTAGATTAATTAATTTGATGAA 1808
Db 1501 AATACTTTTTCAAGCAATGATATGGATGGGCAAGACTTAGATTAATTAATTTGATGAA 1560
QY 1809 GGGCAAAAAGTGGCCCACTAATTAATGATGCTGTAACCTTGATGCTTTCTGCAGAACAG 1868
Db 1561 GGGCAAAAAGTGGCCCACTAATTAATGATGCTGTAACCTTGATGCTTTCTGCAGAACAG 1620
QY 1869 TATCTTCAGACCACTACATAAAGTCTTTTGAAGAAATGTAATGACTCTTAATCGCAA 1928
Db 1621 TATCTTCAGACCACTACATAAAGTCTTTTGAAGAAATGTAATGACTCTTAATCGCAA 1680
QY 1929 ATGAATCAGATAGATATGAAGGCTTAGATGATGGAACATCAATATATATTTCAAT 1988
Db 1681 ATGAATCAGATAGATATGAAGGCTTAGATGATGGAACATCAATATATATTTCAAT 1740
QY 1989 GCAGAACAGAGCTATTTGGGAAGTCATGGTATTAATTAATTTGTTGAAACAGTTGAT 2048
Db 1741 GCAGAACAGAGCTATTTGGGAAGTCATGGTATTAATTAATTTGTTGAAACAGTTGAT 1800
QY 2049 AACAATAACAATAGAAATGGCTTTCTTTAGGAGAAAAAGCACTATTCAGTTCAA 2108
Db 1801 AACAATAACAATAGAAATGGCTTTCTTTAGGAGAAAAAGCACTATTCAGTTCAA 1860
QY 2109 CAAGGGTACCTACAGTAAGTCTGAGATTAACAATCAATTAATCAGTCTCTGATATTAAC 2168
Db 1861 CAAGGGTACCTACAGTAAGTCTGAGATTAACAATCAATTAATCAGTCTCTGATATTAAC 1920
QY 2169 AGTCAATCTGTTGGAGGGCCAGACCTAAGCAATTTGTTAGCCCTTCCATCAAGAACAGG 2228
Db 1921 AGTCAATCTGTTGGAGGGCCAGACCTAAGCAATTTGTTAGCCCTTCCATCAAGAACAGG 1980
QY 2229 AGTTCAAGGACCTGAATAAGCCAGATGTTCCAGATACAATAGAAAGTGAACCCAGACA 2288
Db 1981 AGTTCAAGGACCTGAATAAGCCAGATGTTCCAGATACAATAGAAAGTGAACCCAGACA 2040
QY 2289 GCAGATACCGTTGTTCCAAATCACTTGCTATGATTTCTACAGCTGATCCACAGGTTAGC 2348
Db 2041 GCAGATACCGTTGTTCCAAATCACTTGCTATGATTTCTACAGCTGATCCACAGGTTAGC 2100
QY 2349 TTCAACTTAATTAATGATATAGAAAGTAAATTTCTGAAGTGGATCTAGTTTCGTAAC 2408
Db 2101 TTCAACTTAATTAATGATATAGAAAGTAAATTTCTGAAGTGGATCTAGTTTCGTAAC 2160
QY 2409 GCAATAGAAATCTGTACTGTAACCTGAAACAATTTGAAAGAGCTTGGTTTGGGCCAGAAA 2468
Db 2161 GCAATAGAAATCTGTACTGTAACCTGAAACAATTTGAAAGAGCTTGGTTTGGGCCAGAAA 2220
QY 2469 CAGCCTACTTCGGTTTCTGATTCAGAGCTCCAACTGTATGAATCTGCCAAGTCAAAATTT 2528
Db 2221 CAGCCTACTTCGGTTTCTGATTCAGAGCTCCAACTGTATGAATCTGCCAAGTCAAAATTT 2280
QY 2529 ACTTTTACCAACCGGACACCAATTTGGGAGCATGTGGGAAAGTATTTTGTGGTGTCTGT 2588
Db 2281 ACTTTTACCAACCGGACACCAATTTGGGAGCATGTGGGAAAGTATTTTGTGGTGTCTGT 2340

QY 2589 TGTATAGGAAGTGTAAACTGCAATATCTAGAAAAGGAAGCAAGATATGTAGTCTGC 2648
Db 2341 TGTATAGGAAGTGTAAACTGCAATATCTAGAAAAGGAAGCAAGATATGTAGTCTGC 2400
QY 2649 TATGAAAATATTAGTAAAGCTCAGSCAATTTGAAAGGATGATGAGTCCAACCTGGTTCTAAT 2708
Db 2401 TATGAAAATATTAGTAAAGCTCAGSCAATTTGAAAGGATGATGAGTCCAACCTGGTTCTAAT 2460
QY 2709 CTTAAGTCTAATCATCTGATGAATGTAATCTGAGCTCTCAGCCCTCTCAGAGAACCAACA 2768
Db 2461 CTTAAGTCTAATCATCTGATGAATGTAATCTGAGCTCTCAGCCCTCTCAGAGAACCAACA 2520
QY 2769 TCCAGTATACCTTCCACAGCAACTTTGCGAGTCTCAGCACTTAAACCAACCGAGTGTGAA 2828
Db 2521 TCCAGTATACCTTCCACAGCAACTTTGCGAGTCTCAGCACTTAAACCAACCGAGTGTGAA 2580
QY 2829 GGAATGTTTCCAAAGAACAGAGAGATGATGGTTTTCAGATGGTATATGCCCCAATCGT 2888
Db 2581 GGAATGTTTCCAAAGAACAGAGAGATGATGGTTTTCAGATGGTATATGCCCCAATCGT 2640
QY 2889 GGAATGTCAGATACAAACAAAATTTATCTGGAAGTAAAGATGTTCTGGAAGCTTTAGT 2948
Db 2641 GGAATGTCAGATACAAACAAAATTTATCTGGAAGTAAAGATGTTCTGGAAGCTTTAGT 2700
QY 2949 CCTCTCTCACCTGATGTCCTATGACAGTAAACACAGTGGATCATTCOCATTTCTACTACA 3008
Db 2701 CCTCTCTCACCTGATGTCCTATGACAGTAAACACAGTGGATCATTCOCATTTCTACTACA 2760
QY 3009 GTGGAAGGCCAAAACAATGACAGAGATATTAAGAAGATGAGATTAATCAGAGTCT 3068
Db 2761 GTGGAAGGCCAAAACAATGACAGAGATATTAAGAAGATGAGATTAATCAGAGTCT 2820
QY 3069 ATTTCTCAGGTTCCATCAGTGGAAAAATTTGCTATGAACACAGAGAAATGAGGGTTACCT 3128
Db 2821 ATTTCTCAGGTTCCATCAGTGGAAAAATTTGCTATGAACACAGAGAAATGAGGGTTACCT 2880
QY 3129 ACTCTGGTTCAATTTACCTAGATGATGATGTTTTTCAGAAAACCTGAAGAACCATCTAG 3188
Db 2881 ACTCTGGTTCAATTTACCTAGATGATGTTTTTCAGAAAACCTGAAGAACCATCTAG 2940
QY 3189 CCTACTGGTGTCTTAGTTAAACAGCAATTTACCTATTGCTAGTATTTTCAGATATATAGGTTA 3248
Db 2941 CCTACTGGTGTCTTAGTTAAACAGCAATTTACCTATTGCTAGTATTTTCAGATATATAGGTTA 3000
QY 3249 CTGTGTATTTAAACAAGTATGCTGCAATAAGATTTAGTCTTCTACCTAATGATGAGGAC 3308
Db 3001 CTGTGTATTTAAACAAGTATGCTGCAATAAGATTTAGTCTTCTACCTAATGATGAGGAC 3060
QY 3309 AGTTTCCCCCACTTCTGTTGCATCTGGAGAAAGGATCAGTGCCTGTAGTAGAAGAA 3368
Db 3061 AGTTTCCCCCACTTCTGTTGCATCTGGAGAAAGGATCAGTGCCTGTAGTAGAAGAA 3120
QY 3369 CATCCATCTCATGAGCAGATCAATTTTGTCTTTGAAGGTGAAGGCTTTTCATCCTGTTACA 3428
Db 3121 CATCCATCTCATGAGCAGATCAATTTTGTCTTTGAAGGTGAAGGCTTTTCATCCTGTTACA 3180
QY 3429 TTTGTCCTAAATGCTTAATCTACTCGTGAATGTCAAATTCATATTTTATTCCTCAGACAAA 3488
Db 3181 TTTGTCCTAAATGCTTAATCTACTCGTGAATGTCAAATTCATATTTTATTCCTCAGACAAA 3240
QY 3489 TATTGTTACTTTTCAACCAATGGATTGCGTGGGACAGGCAAGAAATTTATTTCTTA 3548
Db 3241 TATTGTTACTTTTCAACCAATGGATTGCGTGGGACAGGCAAGAAATTTATTTCTTA 3300
QY 3549 TTGTTAAGTTTGCAAAATGAAGATCAATTTCTAAGGACATCTTCAGACTATTTATCACC 3608
Db 3301 TTGTTAAGTTTGCAAAATGAAGATCAATTTCTAAGGACATCTTCAGACTATTTATCACC 3360
QY 3609 ATATATAAGGATGCTCTAAAGGAAAATACATAGAAAACCTTGACAAATATTTACCTTTACT 3668
Db 3361 ATATATAAGGATGCTCTAAAGGAAAATACATAGAAAACCTTGACAAATATTTACCTTTACT 3420
QY 3669 GAGAGTTTCTCAGTAGCAAGGATCAGGAGGATTTCTGTTTATTATACACTACTTTTCAG 3728

[illegible]

Db	4501	CAGCATATCTAAATGATCTTGATAGTCTCTGATACCTGTGTGATCCATGGTGGGACTCC	4560
QY	4809	AACTCTAGTTTACCATTAGAAATAGAAATAGTGTGTTTTTTCATTATAGAACATCTTTTTTAG	4868
Db	4561	AACTCTAGTTTACCATTAGAAATAGAAATAGTGTGTTTTTTCATTATAGAACATCTTTTTTAG	4620
RESULT 5			
BC032227			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
COMMENT			

1335 TCAGCATGACATAAAGATAAATATACAGATGCACTGACTATACATGAAGAAATACAGAA 1394
1505 CAGTGTGTTCTAGTGGGAAACCAATTCAGAGAGAAATGATCTTTTGAACACAGGAAATG 1564
1395 CAGTGTGTTCTAGTGGGAAACCAATTCAGAGAGAAATGATCTTTTGAACACAGGAAATG 1454
1565 TAAAGCATACCTCTTCTAGTGTCTAATTCAGAGAGATGGAAGACAGAAAGATAGATCTCTGA 1624
1455 TAAAGCATACCTCTTCTAGTGTCTAATTCAGAGAGATGGAAGACAGAAAGATAGATCTCTGA 1514
1625 CCAGACAGTAATCAGACAGTGTCTTTGGATGGTGGTGCACACCACTTCTACAGTTGTAGA 1684
1515 CCAGACAGTAATCAGACAGTGTCTTTGGATGGTGGTGCACACCACTTCTACAGTTGTAGA 1574
1685 ATCTCAAGAGGGCTTTCTGGCACTCATGTCCAGAGCTTCTGANTGTGTGAGGTTT 1744
1575 ATCTCAAGAGGGCTTTCTGGCACTCATGTCCAGAGCTTCTGANTGTGTGAGGTTT 1634
1745 TATTAACTCTTTTCAAGCAATGATATGATGGGCAAGACTTAGATTAATTAATTTGA 1804
1635 TATTAACTCTTTTCAAGCAATGATATGATGGGCAAGACTTAGATTAATTAATTTGA 1694
1805 TGAAGGCGCAAAAGTGGCCCACTAATTAATGATGCTGAACCTTGATGCTTTCTGACAGA 1864
1695 TGAAGGCGCAAAAGTGGCCCACTAATTAATGATGCTGAACCTTGATGCTTTCTGACAGA 1754
1865 ACAGTATCTTCAGACCACTACATTAAGTCTTTTGAAGAAATGTAATGATCTCAATC 1924
1755 ACAGTATCTTCAGACCACTACATTAAGTCTTTTGAAGAAATGTAATGATCTCAATC 1814
1925 GCAATGAATCAGATAGATATGAAGGCTTAGATGATGCAACCATCAATTAATTAATTT 1984
1815 GCAATGAATCAGATAGATATGAAGGCTTAGATGATGCAACCATCAATTAATTAATTT 1874
1985 CAATCAGAGCAGAGCTATTTGGGAAAGTCATGGTATTAATTAATTAATTTGTGAAACAGT 2044
1875 CAATCAGAGCAGAGCTATTTGGGAAAGTCATGGTATTAATTAATTAATTTGTGAAACAGT 1934
2045 TGATAACCAAAATCAATAGAAAATGGCTTTCTTTAGGAGAAAGCACTATTTCCAGT 2104
1935 TGATAACCAAAATCAATAGAAAATGGCTTTCTTTAGGAGAAAGCACTATTTCCAGT 1994
2105 TCAACAGGGTTACCTACCACTAGTCTGAGATTAACAAATCAATTAATCACTCTGATAT 2164
1995 TCAACAGGGTTACCTACCACTAGTCTGAGATTAACAAATCAATTAATCACTCTGATAT 2054
2165 TAACAGTCAATCTGTGGAGGGCCAGACCTTAAGCAATTTGTAGCTTCCATCAAGAAC 2224
2055 TAACAGTCAATCTGTGTGGAGGGCCAGACCTTAAGCAATTTGTAGCTTCCATCAAGAAC 2114
2225 AAGGAGTTCAAAGGACCTTGAATAAGCCAGATGTTCCAGATACATAAGAAAGTGAACCCAG 2284
2115 AAGGAGTTCAAAGGACCTTGAATAAGCCAGATGTTCCAGATACATAAGAAAGTGAACCCAG 2174
2285 CACAGCAGATACGTTGTTCCAACTACCTGTGCTATAGATTTACAGCTGATCCACAGGT 2344
2175 CACAGCAGATACGTTGTTCCAACTACCTGTGCTATAGATTTACAGCTGATCCACAGGT 2234
2345 TAGCTTCAACTCAATTAATGATAGAAAGTAAATTTCTGAAGGTGGATCTAGTTTCGT 2404
2235 TAGCTTCAACTCAATTAATGATAGAAAGTAAATTTCTGAAGGTGGATCTAGTTTCGT 2294
2405 AACTGCAAAATGAAGATCTGTACCTGAAACCACTTGCAGAAAGAGGCTGTGTTTGGGCCA 2464
2295 AACTGCAAAATGAAGATCTGTACCTGAAACCACTTGCAGAAAGAGGCTGTGTTTGGGCCA 2354
2465 GAAACAGCCTACTTGGGTTCTGATTCAGAGGCTCCAACTGTATGAAGTCCCAAGTCAA 2524
2355 GAAACAGCCTACTTGGGTTCTGATTCAGAGGCTCCAACTGTATGAAGTCCCAAGTCAA 2414
2525 ATTTACTTTTACCAAAACGGGACACCAATTCGAGAGCATGTGGGAAAGTATTTTGTGGTGT 2584

2415 ATTTACTTTTACCAAAACGGGACACCAATTCGCGAGCATGTGGGAAAGTATTTTGTGGTGT 2474
2585 CTGTTCTAATAGGAAGTGTAAACTGCAATATCTAGAAAAAGGAGCAAGAGTATGTGTAGT 2644
2475 CTGTTCTAATAGGAAGTGTAAACTGCAATATCTAGAAAAAGGAGCAAGAGTATGTGTAGT 2534
2645 CTGCTATGAAAATCTATTAGTAAAGCTCAGGCAATTTGAAAGATGATGAGTCCAACTGGTTC 2704
2535 CTGCTATGAAAATCTATTAGTAAAGCTCAGGCAATTTGAAAGATGATGAGTCCAACTGGTTC 2594
2705 TAATCTTAAGTCTAACTATTCTGATGAATGTACTCTGCTCCAGCTTCCCTCAGGAGAACCA 2764
2595 TAATCTTAAGTCTTAATCTATCTGATGAATGTACTCTGCTCCAGCTTCCCTCAGGAGAACCA 2654
2765 AACATCCAGTATACCTTCCAGGCAACTTTTGCAGCTCTCAGCACTTAAACACCAAGGTGT 2824
2655 AACATCCAGTATACCTTCCAGGCAACTTTTGCAGCTCTCAGCACTTAAACACCAAGGTGT 2714
2825 TGAAGCACTATGTTCCAAAGAAACAGAGAGATGTGTTTGCAGATGATATTCGCCAA 2884
2715 TGAAGCACTATGTTCCAAAGAAACAGAGAGATGTGTTTGCAGATGATATTCGCCAA 2774
2885 TGGTGAAGTTGCAGATACAAACAAAATTTATCTCTGGAAGTAAAGATGTTCTTGAAGACTT 2944
2775 TGGTGAAGTTGCAGATACAAACAAAATTTATCTCTGGAAGTAAAGATGTTCTTGAAGACTT 2834
2945 TAGTCTCTCTCAGCTGATGTGCTTATGACAGTAAACACAGTGGATCATTCCTATCTAC 3004
2835 TAGTCTCTCTCAGCTGATGTGCTTATGACAGTAAACACAGTGGATCATTCCTATCTAC 2894
3005 TACAGTGGAAAGCCAAACAAATGACAGAGATATTAACAGAAATGAGATAATTCAGAG 3064
2895 TACAGTGGAAAGCCAAACAAATGACAGAGATATTAACAGAAATGAGATAATTCAGAG 2954
3065 TCTATTTCTCAGGTTCCATCAGTGGAAAAATTTCTATGAAACACAGGAAATAGGGGTT 3124
2955 TCTATTTCTCAGGTTCCATCAGTGGAAAAATTTCTATGAAACACAGGAAATAGGGGTT 3014
3125 ACCTACTCTCTGTTCAATTTACACTAGATGATGATGTTTTCAGAGAACTGAGAACCATC 3184
3015 ACCTACTCTCTGTTCAATTTACACTAGATGATGATGTTTTCAGAGAACTGAGAACCATC 3074
3185 TAGTCTCTCTGTTCTTAGTTTAAACAGCAATTTACCTATTTCTAGTATTTTCCAGATTATAG 3244
3075 TAGTCTCTCTGTTCTTAGTTTAAACAGCAATTTACCTATTTCTAGTATTTTCCAGATTATAG 3134
3245 GTTACTGTGTGATATTAACAGATGATCTGCAATGATGATGTTTCTTACCTAATGATGA 3304
3135 GTTACTGTGTGATATTAACAGATGATCTGCAATGATGATGTTTCTTACCTAATGATGA 3194
3305 GGACAGTTTGGCCCACTTCTGGTTGCACTTGGAGAAAGGATCAGTGCCTGTAGTAGA 3364
3195 GGACAGTTTGGCCCACTTCTGGTTGCACTTGGAGAAAGGATCAGTGCCTGTAGTAGA 3254
3365 AGAACATCCATCTCATGACAGATCAATTTGCTTCTTGAAGGTGAGGCTTTCATCTGT 3424
3255 AGAACATCCATCTCATGACAGATCAATTTGCTTCTTGAAGGTGAGGCTTTCATCTGT 3314
3425 TACATTTGTCTTAAATGCTAATCTACTGTAATGATGTAATTTTATTTTCTCTCAGA 3484
3315 TACATTTGTCTTAAATGCTAATCTACTGTAATGATGTAATTTTATTTTCTCTCAGA 3374
3485 CAAATATTGGTATCTTTTCAACCAATGGATTCGATGCTGGACAGGAGGAGAAATATTAT 3544
3375 CAAATATTGGTATCTTTTCAACCAATGGATTCGATGCTGGACAGGAGGAGAAATATTAT 3434
3545 TCTATTGTTATGTTTGGCCAAATGAAGTACTATTTCTTAAAGGACATCTTCAGACTATTAT 3604
3435 TCTATTGTTATGTTTGGCCAAATGAAGTACTATTTCTTAAAGGACATCTTCAGACTATTAT 3494
3605 CACCATATATAAGGATGCTCTTAAAGGAGAAATACATAGAAACTTTGAGCAATATTACCTT 3664
3495 CACCATATATAAGGATGCTCTTAAAGGAGAAATACATAGAAACTTTGAGCAATATTACCTT 3554

QY 3665 TACTGAGAGTTTCTAGTAGCAGGATCAGGAGGATTCCTGTTTATTACACCTACTTT 3724
Db 3555 TACTGAGAGTTTCTAGTAGCAGGATCAGGAGGATTCCTGTTTATTACACCTACTTT 3614
QY 3725 TCAGAACTTGATGATCTCTCATTACCAAGTAATCCTTTTCTTTGTTGGAAATCTTATCCA 3784
Db 3615 TCAGAACTTGATGATCTCTCATTACCAAGTAATCCTTTTCTTTGTTGGAAATCTTATCCA 3674
QY 3785 GAAGCTTGAGATCCCTGGGCAAGGTTTTCCTATGCGTTTAAATGTTGAGATGGGTGC 3844
Db 3675 GAAGCTTGAGATCCCTGGGCAAGGTTTTCCTATGCGTTTAAATGTTGAGATGGGTGC 3734
QY 3845 AGAATATAAG 3855
Db 3735 AGAATATAAG 3745

RESULT 6
AK127003
LOCUS 3793 bp mRNA linear PRI 09-SEP-2003
DEFINITION Homo sapiens cDNA FLJ45058 fis, clone BRAWH3023172, moderately similar to Mothers against decapentaplegic homolog interacting protein.
ACCESSION AK127003
VERSION AK127003.1 GI:34533719
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Wakami, B., Sugiyama, A., Takemoto, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3793)
AUTHORS Isogai, T. and Yamamoto, J.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.
FEATURES
source
1.3793
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="BRAWH3023172"
/tissue_type="brain"
/clone_lib="BRAWH3"
/note="cloning vector: pME18SFL3"

ORIGIN
Query Match 51.6%; Score 3418.8; DB 9; Length 3793;
Best Local Similarity 93.4%; Pred No. 0;
Matches 3709; Conservative 0; Mismatches 2; Indels 259; Gaps 3;
QY 2663 TAAAGCTCAGGCATTTGAAAGGATGATGAGTCCAACTGGTCTTAATCTTAAGTCTAATCA 2722
Db 70 TATAGCTCAGGCATTTGAAAGGATGATGAGTCCAACTGGTCTTAATCTTAAGTCTAATCA 129

2723 TTCTGATGAATGTAATACTGCTCCAGCCCTCTCAGGAGAACCAAAATCCAGTATACCTTC 2782
Db 130 TTCTGATGAATGTAATACTGCTCCAGCCCTCTCAGGAGAACCAAAATCCAGTATACCTTC 189
QY 2783 ACCAGCAACTTTGGCAGTCTCAGCACTTTAAACCAACAGGTGTTGAAGCACTATGTTCAA 2842
Db 190 ACCAGCAACTTTGGCAGTCTCAGCACTTTAAACCAACAGGTGTTGAAGCACTATGTTCAA 249
QY 2843 AGAAGAGAGAGATGTTGGTTCAGATGGTATATGGCCATGAGTGAAGTTCAGATAC 2902
Db 250 AGAAGAGAGAGATGTTGGTTCAGATGGTATATGGCCATGAGTGAAGTTCAGATAC 309
QY 2903 AACAAATTTATCATCTGGAAGTAAAGATGTTCTCAAGACTTTTACTCTCTCTCACTGA 2962
Db 310 AACAAATTTATCATCTGGAAGTAAAGATGTTCTCAAGACTTTTACTCTCTCTCACTGA 369
QY 2963 TGTGCTATGACAGATAACACAGTGGATCTCCCATTTCTACTACAGTGGAAAGCCAAA 3022
Db 370 TGTGCTAT 378
QY 3023 CAATGAGACAGGAGATATTACAAGAAATGAGATAAATTCAGAGTCTCTATTCTCAGGTTC 3082
Db 379 378
QY 3083 ATCAGTGGAAAAATTTCTATGACACAGGAAATGAGGGTTTACCTCTTCTGGTTCAAT 3142
Db 379 378
QY 3143 TACACTAGATGATGATGTTTTCAGAACTGAGAACCATCTAGTCTCTACTGTTGCTTT 3202
Db 379 378
QY 3203 AGTTAACGCAATTTACCTATTGCTAGTATTTTTCAGATATAGGTTTACTGTGTGATATAA 3262
Db 379 412
QY 3263 CAAATGATCTGCAATAAGATTTAGTCTTCTACTAATGATGAGGACAGTTTCCGCCACT 3322
Db 413 CAAATGATCTGCAATAAGATTTAGTCTTCTACTAATGATGAGGACAGTTTCCGCCACT 472
QY 3323 TCTGTTGATCTGGAGAAAGGATCAGTGGCTGTAGTAGAAGAACATCCATCTCATCA 3382
Db 473 TCTGTTGATCTGGAGAAAGGATCAGTGGCTGTAGTAGAAGAACATCCATCTCATCA 532
QY 3383 GCAGATCAATTTGCTTTGAAAGGTGAAGCTTTTCACTCTGTTTACATTTGTCTTAAATGC 3442
Db 533 GCAGATCAATTTGCTTTGAAAGGTGAAGCTTTTCACTCTGTTTACATTTGTCTTAAATGC 592
QY 3443 TAATCTACTCTGTAATGTCATATTTTATTTCTCCTCAGACAAATATTTGTTACTTTTC 3502
Db 593 TAATCTACTCTGTAATGTCATATTTTATTTCTCCTCAGACAAATATTTGTTACTTTTC 652
QY 3503 AACCAATGATTTGCTGCTTGGGACAGGCAAGAAATTTATTTCTATTGTTTGTGCT 3562
Db 653 AACCAATGATTTGCTGCTTGGGACAGGCAAGAAATTTATTTCTATTGTTTGTGCT 712
QY 3563 AATGAGATCAATTTCTTAAAGACATCTTCAGACTATTTTATCACCATATATAAGATGC 3622
Db 713 AATGAGATCAATTTCTTAAAGACATCTTCAGACTATTTTATCACCATATATAAGATGC 772
QY 3623 TCTAAAGGAAATATACATAGAAAACCTTGACAAATATTACTTTTACTGAGAGTTTCTCAG 3682
Db 773 TCTAAAGGAAATATACATAGAAAACCTTGACAAATATTACTTTTACTGAGAGTTTCTCAG 832
QY 3683 TAGCAAGGATCAGGAGATTTCTGTTTATTACCTACTTTTTCAGAAAATTTGATGATCT 3742
Db 833 TAGCAAGGATCAGGAGATTTCTGTTTATTACCTACTTTTTCAGAAAATTTGATGATCT 892
QY 3743 CTCAATACCAAGTAATCCCTTTTCTTTGTTGAAATTTCTTATCCAGAACTTTGAGATCCCTG 3802
Db 893 CTCAATACCAAGTAATCCCTTTTCTTTGTTGAAATTTCTTATCCAGAACTTTGAGAT 951

Db	3111	TTACAGAGCTTCTCTTCTTACTTCAACACAGCAAAAAGTGGGGGCATATTGTAGTCTGT	3170
Qy	6023	TCATTTAAGTTATGTAATAAATTAATCATATTATTTGATGCTTTAAACATTTCTCATGCT	6082
Db	3171	TCATTTAAGTTATGTAATAAATTAATCATATTATTTGATGCTTTAAACATTTCTCATGCT	3230
Qy	6083	AAATATATGTTTTGTATCAAAAACACTCATATATTTCAAGAAAAAGAAATATGTTAAAT	6142
Db	3231	AAATATATGTTTTGTATCAAAAACACTCATATATTTCAAGAAAAAGAAATATGTTAAAT	3290
Qy	6143	AGCCTCTTTTAAAGAAAAATTTATGAAGCATCTCACTTGAAGATCAAGTCAAAAGTTA	6202
Db	3291	AGCCTCTTTTAAAGAAAAATTTATGAAGCATCTCACTTGAAGATCAAGTCAAAAGTTA	3350
Qy	6203	TAAGTCAGGATCTGAGGCTCAAGCTAGGAGAGACTGAGAAATTTTAAATCAAGTTGGGCAT	6262
Db	3351	TAAGTCAGGATCTGAGGCTCAAGCTAGGAGAGACTGAGAAATTTTAAATCAAGTTGGGCAT	3410
Qy	6263	ATAGTTTGGACTGAATCACAATCTGTAGTACTTACCCAAAGCAATTTGGAGGAGATATC	6322
Db	3411	ATAGTTTGGACTGAATCACAATCTGTAGTACTTACCCAAAGCAATTTGGAGGAGATATC	3470
Qy	6323	AGCCTCTGGAAGTAGTCTCTCTGGAACAATGTAAGTCTCCAGATATTCATATAAT	6382
Db	3471	AGCCTCTGGAAGTAGTCTCTCTGGAACAATGTAAGTCTCCAGATATTCATATAAT	3530
Qy	6383	GGCAACCTGTTATATTTGTAATTTTGAATTTTATGAATGCTGAAGATGAAAACAATTCGATA	6442
Db	3531	GGCAACCTGTTATATTTGTAATTTTGAATTTTATGAATGCTGAAGATGAAAACAATTCGATA	3590
Qy	6443	TCAAACCAATTTATGTTTCTTAAATATAGTGTATGTTATGCTGCAATGAAGTAATGAA	6502
Db	3591	TCAAACCAATTTATGTTTCTTAAATATAGTGTATGTTATGCTGCAATGAAGTAATGAA	3650
Qy	6503	CAGTCTTAAATACCAATGAGTGAAGGCTGTTCCATGATGGACAGCTTGGATTGT	6562
Db	3651	CAGTCTTAAATACCAATGAGTGAAGGCTGTTCCATGATGGACAGCTTGGATTGT	3710
Qy	6563	TTTCATAAAATCTCTACATTCATCAATCAATTTGAAATTTATGTCCTGAAGTTGAGGCAC	6622
Db	3711	TTTCATAAAATCTCTACATTCATCAATCAATTTGAAATTTATGTCCTGAAGTTGAGGCAC	3770
Qy	6623	ATTTTGAAGT 6632	
Db	3771	ATTTTGAAGT 3780	
RESULT 7			
AK122241	LOCUS		
DEFINITION	Mus musculus mRNA for mKIAA0305 protein.		
ACCESSION	AK122241		
VERSION	AK122241.1 GI:28972138		
KEYWORDS	FLI CDNA.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Aizawa,H., Yuasa,S., Nakajima,D., Nagase,T., Ohara,O. and Koga,H.		
TITLE	Prediction of the coding sequences of mouse homologues of KIAA gene. II. The complete nucleotide sequences of 400 mouse KIAA-homologous cDNAs identified by screening of terminal sequences of cDNA clones randomly sampled from size-fractionated libraries		
JOURNAL	DNA Res. 10, 35-48 (2003)		
REFERENCE	2 (bases 1 to 5438)		
AUTHORS	Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research Institute, Laboratory for Genome Informatics; 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)		

COMMENT	The CREATE program supported by Japan science and technology corporation; cDNA full insert sequencing; Kazusa DNA Research Institute; cDNA library construction, clone selection and 5'- & 3'-end one pass sequencing.		
FEATURES	Location/Qualifiers		
source	1..5438		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10090"		
	/clone="mbg08215"		
	/tissue_type="brain"		
	/dev_stage="adult"		
	/note="vector:modified pBC SK+"		
gene	1..5438		
	/gene="mKIAA0305"		
CDS	<109..4719		
	/note="CDS is predicted by in silico analysis. Start codon is not identified."		
	/evidence="not experimental"		
	/product="mKIAA0305 protein"		
	/protein_id="BAC65523.1"		
	/db_xref="GI:28972139"		
	/translation="VLSKPTGRMDSYFKAAYSGLDKLLDDPEONPKODYLPDAYAFN QVSSESAGQLALLSKDQCISTCASSEACCEDANETLEGIHEGLTSRPNEKNV AGLDLSVSDASTDEIQPCWRRCKFVCDLISDMGLVHATNSEEDIKQLPDPK ASADLTLDSSVSEALTVSSVDCGNAVREEQNNINAGIKRNDISKEJGVKVDMA LPDCKYRNTENLKXII:SNLETVDFDMPVLMKEQSEMGTNDKISPKYKPCLLK DQCLAEKVAVANNTCELBEGGSSNTIAMPKLPENEGISPSDPASKDENKFLPFD PLOENTSVFMKOTVKEDSRNLDKNNDIIVHSGDDVPPLSLCLSLSGSLGKSLIHN NEHSLLPNSESOGNDNAVIIHEIOKSDVLDGETDLSKCTCSI:FLQPVNKKGG QKVEVEVMISGESLESDEDSASAAASPVASVPEAPGCEGTFFPSDWMDGQ ELDFYFNIDESMRSGILISDAELDAFLKEQCSLNSNTMSAGENVDSQLQMQITWKL HDENAGDIYFAEAGAGENGVCNETSKENTENGLSIGEKGAIPERELSAQCP DIRDLVPISIKTAAGGAPKQLSLPFGTRSKELNKPDVVDVPESEPTANATV STCSADHLPDSOVSPNSNVID:ESNFEDGSSFTANKDSLKENKESILVQKQFTW VPSAPCMNCQVKTFTKRHRHCGKGVFCVCCNCKLOYLEKEARVCVICYE TINKQAQERWMSPGSCLKNHNECATDPOLOTOJSTSPSTPLTISALKQNV EGPCEKQKFWFADGILFNGEVADITKLSSSKRSDPSPVLPVPMINKVDRTHS PTERKPNGLDIIRSEISQSPCTHAPVERLPGNTGEGPMPGPFLEDDVFDSE BESTPTVPSNGSLPVASISDYRLICSVAKVCNNISLLPDDDIGLPLLATSGEDGS VPVQVERPSHEQIILLLEGEPFPATFVLNANLVNLVLYSSYKESWFFSNGHLG GOAEIIVLLQCLPNEDTPKQIFRLFITIKDALKKYIENLDNLTFTFESFLNKDHG GFLFITPQNLGDPVPSRPFGLIGLQKLEI:PWAKVPMRLMLRLGAERYKAPFL TSVGRKPLFGEIGHTIKNLVLDNRYOYTLHNDQLIHMEMGSKCI:PRKYSDV MKVTHSSNEHVISIGASSTEDSHLVQSGDVYQTCANSATGPRKVTGASVUVFN GALTSSGSLAKSIVEDGLMVQITPETMEGRLLAREKQDFRIGCKKVDVLDREYV DICWDSSEERKNKGVISVDGMSVEGFPSEKIGETDFETEEXTVACTEVFFLKDDQ ISILSSSYQFAKEIAVACSAALCPHLRTLKSNRNKIGLRVSDIDTMVFEQACGGQL LPQHYLNDLDSALIPVINGGTSNSSLPLEIEIAFFILENLSE"		
ORIGIN			
Query Match	50.3%;	Score 3337.4;	DB 10; Length 5438;
Best Local Similarity	78.9%;	Pred. No. 0;	
Matches 4303;	Conservative 0;	Mismatches 1021;	Indels 133; Gaps 22;
Qy	208	AGGCATCAAGGAATTAATAATCTGAATAAGTCTGCAGGTAGGATGGACAGTATTTTAAAG	267
Db	92	AGCATAGAGGAATGAGTTCAGTAACCCACAGGAGGATGGACAGTACTTTTAAAG	151
Qy	268	CAGCTGTCAGTGACTTGGACAAACTCCTTGTATGATTTTGAACAGAACCCAGATGAACAAG	327
Db	152	CAGCTGTCAGTGGCTTGGACAAACTTCTTCATGATTTTGGAGCAGATCCAGATAAGCAAG	211
Qy	328	ATTATCTCGAGATGTACAAAATGCATATGATTTCAACCACTGCTCAGTCTTCTTCAGAGT	387
Db	212	ATTATCTCCAG-----ATGCATATGCTTTTAAACCACTGCTCAGTCTTCTTCAGAGT	262
Qy	388	TGGCTTCTCCTCACAGCAAGTCTTCAATTCCTCCAAAAGCAAGAGTCCGTTAATAGTTG	447
Db	263	CGGCTCTCCACAGCTAGCT---TTACTGTCAAAGGACCAAGATGCATCAGTACTTGTG	319

Dd 2435 GTTGTACAGGAATGTAAAGCTTCACTATCTAGAGAGAGCAAGATGTGTAACT 2494
Qy 2647 GCTATGAACCTATPAGTAAAGCTCAGGCATTTGAAAGATGATGAGTCCAACTGGTCTTA 2706
Dd 2495 GCTATGAGACTATAAATAAAGCTCAGGCATTTGAAAGATGATGAGTCCAACTGGTCTTA 2554
Qy 2707 ATCTTAAGTCTCAATCTATGATGAATGTACTAGTCCAGCTCTCCTCAGGAGAACCAA 2766
Dd 2555 GTCTTAAGTCTCAATCTATGATGAATGTACTAGTCCAGCTCTCCTCAGGAGAACCAA 2614
Qy 2767 CATCCAGTATACCTTCAACAGCACTTTGCGAGTCTCAGCACTTAAACAACAGGTTG 2826
Dd 2615 CATCCAGTATACCTTCAACAGCACTTTGCGAGTCTCAGCACTTAAACAACAGGTTG 2674
Qy 2827 AAGCACTATGTTCCAAAGACAGAGAGATGTTGCGAGTCTCAGCACTTAAACAACAGGTTG 2886
Dd 2675 AAGCACTATGTTCCAAAGACAGAGAGATGTTGCGAGTCTCAGCACTTAAACAACAGGTTG 2734
Qy 2887 GTGAAGTTGCGAGTATACCAACAAATTTATCATCTGGAAGTAAAGATGTTCTGAGACCTTTA 2946
Dd 2735 GTGAAGTTGCGAGTATACCAACAAATTTATCATCTGGAAGTAAAGATGTTCTGAGACCTTTA 2794
Qy 2947 GTCCTCTCTCACTGATGTCGTATGACAGTAAACAGTGGATCATTCCTCCATTTCTACTA 3006
Dd 2795 GTCCTCTCTCACTGATGTCGTATGACAGTAAACAGTGGATCATTCCTCCATTTCTACTA 2854
Qy 3007 CAGTGGAAAGCCCAACAAATGAGACAGAGATATTACAGAAATGAGATAATTCAGAGTC 3066
Dd 2855 CAGTGGAAAGCCCAACAAATGAGACAGAGATATTACAGAAATGAGATAATTCAGAGTC 2914
Qy 3067 CTATTTCTCAGGTTCCATCAGTGGAAATAATGTCATGAAACAGAGAAATGAGGGTTAC 3126
Dd 2915 CAACCTGTCATCTGCACTCTGCGAAAGATGTCGCGGGAACACAGGACGGAAGGTTGC 2974
Qy 3127 CTACTCTGCTTCACTTACACTAGATGATGTTTTCAGAACTGGAAGACCACTTA 3186
Dd 2975 CCATGCTGCGCTTTCACACTAGAGATGATGTTTTCAGAACTGGAAGACCACTTA 3034
Qy 3187 GTCCTACTGCTGCTTCACTTAAACCAATTTACCTATGCTAGTATTTCAGATTATAGGT 3246
Dd 3035 CTCCTACTGTTGCTCCTGCTTAATAGTGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3094
Qy 3247 TACTGTGTGATATTAAACAGATGTCGCAATAGATGATGCTTCTACCTTAATGATGAGG 3306
Dd 3095 TANTGTGTGATGTTGCGAAATGTCGCAACACATTTAGTCTTCTACCAAGACGATGACA 3154
Qy 3307 ACAGTTTGCCTCCCTCTGCTTGCATCTGAGAAAGGATCAGTCCCTGCTAGTAGAAG 3366
Dd 3155 TTGTTTGCCTCCCTCTGCGCACATCTGGAGAGATGATCAGTCCCTGTTGTGCAAG 3214
Qy 3367 AACATCCATCTCATGAGCAGATCAATTTGCTTCTTGAAGTGAAGGCTTTTCATCCTGTTA 3426
Dd 3215 AACGTCCTCTCATGAGCAGATCAATTTTACTTCTTGAAGGGAAGGATTTTCTCCTGCTA 3274
Qy 3427 CATTTGCTTAAAGTCTAACTCTCTGATGATGCAATTCATATTTTATTTCTTCAGACA 3486
Dd 3275 CATTTGTTCTAAATGCAAACTCTCTGATGATGCAATTTGCTGCTTATTTCTTCAGAGA 3334
Qy 3487 AATATTGCTATTTTCAACCAATGATGATGCTTGGGACAGGCAAGAAATTTATTATTC 3546
Dd 3335 AGTATTGCTATTTCTCAACCAATGATGATGCTTGGGACAGGCAAGAAATTTATTATTC 3394
Qy 3547 TATTGTTATGTTTGCAGAAATGAGATCTATCTTAAGGACATCTTCAGACTATTATCA 3606
Dd 3395 TGTTGCAATGTTTGCAGAAATGAGATCTATCTTAAGGACATCTTCAGACTATTATCA 3454
Qy 3607 CCATATAGGATGCTTAAAGGAAATACATAGAAACTTGGACAAATATTACCTTTA 3666
Dd 3455 CCATATAGGATGCTTAAAGGAAATACATAGAAATTTGGACAAATATTACCTTTA 3514
Qy 3667 CTGAGAGTTTCTCAGTAGCAAGGATCAGGAGGATTTCTGTTTATTATACCTACTTTTC 3726
Dd 3515 CTGAGAGTTTCTCAACAGCAAGGATCATGAGGTTTCTGTTTATTATACCTACTTTTC 3574

Qy 3727 AGAACTTGTATGATCTCTCATTAACAAGTAACTCTTTTCTTTGTGGAATTTCTTATCCAGA 3786
Dd 3575 AGAACTTGTATGATCTCTCCCGTGCACAGAGTCTCTTTCTCTGCGGCATTTCTTATCCAGA 3634
Qy 3787 ACCTTCAAGATTTCCCTGGGCAAAAGGTTTTTCTCTATGCTTTTAAATGTTGAGATTTGGGTGAG 3846
Dd 3635 AGCTGGAGATTTCCCTGGGCAAAAGGTTTTTCTCTATGCGGCTAATGTTGCGATTGGGAGCAG 3694
Qy 3847 AATATAAGCATATCTCTGCTCTCTAACAAGCATCAGAGGCGGAAAACTCTCTTTTGGAG 3906
Dd 3695 AATACAAAAGCATPACCTGCTCTCTCTAACAAGTGTCTGAGGCGAGGAACTCTCTTTTGGAG 3754
Qy 3907 AAATAGGACACACTATTATGAACCTTTACTTGTGTGACCTTCGAAATTTACCAGTATACCTTGC 3966
Dd 3755 AAATAGGACACACTATTATGAACCTTTACTTGTGTGACCTTCGAAATTTACCAGTATACCTTGC 3814
Qy 3967 ATAATATAGATCAACTGTTGATTCATATGAAATGGGAAAAAGCTGCAATAAAAAATACAC 4026
Dd 3815 ATAACATCATCAGCTAATGATTCATATGAAATGGGAAAAAGCTGCAATAAAAAATACAC 3874
Qy 4027 GGAAGAGTACAGTGTATGTAATGAAAGTAAATTTCTTCCATCAGCATGTCATTAGCA 4086
Dd 3875 GAAAGAGTACAGTGTATGTAATGAAAGTAAATTTCTTAAACGAGCACGTCATCAGTA 3934
Qy 4087 TTGGAGCAAGTTTTCAGTACAGAGCAGATTTCTCATCTAGTCTGTATACAGAAATGATGAA 4146
Dd 3935 TTGGAGCTAGTTTATGTAAGAGCAGATTTCTCATCTGCTGTGTACAGAGTGTATGAG 3994
Qy 4147 TTTATGAAACACAGAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4206
Dd 3995 TTTATCAGACTCAGGCAACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4054
Qy 4207 TTGTGCTATTCATGAGCTCTAAAAACATCTTCAGGATTTCTTCTAAGTCCAGCATAG 4266
Dd 4055 TTGTGCTGTTTATGCTGCTCTGAAACATCTTCAGGATTTCTTCCAAATCCAGCATAG 4114
Qy 4267 TTGAAGATGGCTTAAATGTAAGAAATTAATCAGAGACCATGATGATGCTGCTGCTGCTGCTGCT 4326
Dd 4115 TCGAAGATGGCTTAAATGTAAGAAATTAATCAGAGACCATGATGATGCTGCTGCTGCTGCTGCT 4174
Qy 4327 TACGAGAACAGAAAGATTTTAAATTTACATGTTGGGAAAGTTTCTGATGAGTACCTGAGAG 4386
Dd 4175 TACGAGAACAGAAAGATTTTAAATTTAGAAATCCAGTGTGGAAGTGTGCTGCTGCTGCTGCTGCT 4234
Qy 4387 AATAAGTGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4446
Dd 4235 AGTATGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4294
Qy 4447 CAGTGTGGAATATCATTAAGGATTTTCCAAAGTAAATTAATTAATTAATTAATTAATTAATTAAT 4506
Dd 4295 CAGTGTGGAATATCATTAAGGATTTTCCAAAGTAAATTAATTAATTAATTAATTAATTAATTAAT 4354
Qy 4507 TTGAAACCCGATCAGAGATTTTAAATGTAAGGATTTTCTTACTTTCTTAAAGGACCAAG 4566
Dd 4355 TTGAAACTCAGGAGAGATCTGTAAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4414
Qy 4567 ATTATCTATTATTAATCACTTCTTATCAGTTTGGAAAGAAATAGCCATGCTGCTGCTGCTGCT 4626
Dd 4415 ATATATCTATTATTAATCACTTCTTATCAGTTTGGAAAGGATTTGCTGCTGCTGCTGCTGCT 4474
Qy 4627 CTGCGCT 4686
Dd 4475 CTGCGCT 4534
Qy 4687 GAGTTTCCATTAATCACTGATGATGTTGAAATTTTCCAGGAGGATTTGAGGCCCACTTTCTGCT 4746
Dd 4535 GAGTGTCCATTAATCACTGATGATGTTGAAATTTTCCAGGAGGATTTGAGGCCCACTTTCTGCT 4594
Qy 4747 CTCAGCATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4806
Dd 4595 CTCAGCATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4654

QY 4807 CCAACTCTAGTTTACCATTAGAAATAGATAGTGTGTTTTCATTTATAGAACATCTTTTT 4866
 Db |||||
 QY 4655 CCAACTCTAGTCTGCCACTAGAAATAGATAGTGTGTTTTCATTTATAGAACATCTTTCTG 4714
 Db |||||
 QY 4867 AGTGAAAGAAATGTGCATATATACATATTTGAACTTAATTTGTTTAAACTTAATCCACAC 4826
 Db |||||
 QY 4715 AGTGAAACCGGTACTGTGTGTATCTTGTGAACTGATCGGTTTAAACGAAATCCACAC 4774
 Db |||||
 QY 4927 TAAAGCTCAAAATGCCACAAACACTAAAGATATAATATGCTGATTTTGAACACATAA 4986
 Db |||||
 QY 4775 TAAACGGAATGCCACCAACTAA-----AGTTTGATATTTGAAATATAA 4824
 Db |||||
 QY 4987 GCTTGTCTTTAGGACGAAATGATCTTTTCAATCATATAGCAAAATTTAAATATCTA 5046
 Db |||||
 QY 4825 -TTTGTCTTTTAGTAGGAATGATTTTAAATATCATATAGCAGATTTAAGTTCTCTA 4883
 Db |||||
 QY 5047 AAAATTTAAGATCCATCTCTGTTAGCTTTTCAATTAATTTAAGTACTTAAAGACAA 5106
 Db |||||
 QY 4884 AAAATTTAAG-GAATCATCATTTTATAGCTTTATGCTTATATAGAACTAAGAGG--A 4940
 Db |||||
 QY 5107 AGGATTTCTTTTAAAGAAATTTATAGCAATTTACTGTGTTTAAATGCTAAGCCAAAGTA 5166
 Db |||||
 QY 4941 AATATTTCTAATCAATTTACAGTATTTTACTGTGTACCTTAACTGTTAAGCCAAAGCA 5000
 Db |||||
 QY 5167 TCTGCACTTAGTATACCTCTTTATGCCAATTAATTTAATGAAGCTCTTTTCAGAT 5226
 Db |||||
 QY 5001 TCTGCGCTTAGCCCTACTCTCATAGATGACGA----- 5033
 Db |||||
 QY 5227 GTAACCTTATGAAGAAATATCTGTTTGTATATGCCAGTTAGAAATCTGTTTCTTAA 5286
 Db |||||
 QY 5034 -----GAATGAAGAGAGACCTGGTGTGTGATGCTAGTGAGACACTGGTTT-TGA 5086
 Db |||||
 QY 5287 AGTGTGCAAAATGTATTTCTAGTGAGCAAAACACAGTTTGTAGGCTCTTGTAGACTTAAAT 5346
 Db |||||
 QY 5087 AGTCTGTTTAAATATGATCTCAGTAGAAGAACAGCTTTCAAGATTTCTAGACTTAAAT 5446
 Db |||||
 QY 5347 TCTTTGATATAAATGATAACTTATTTGTATATTAATTTGGAGTGAGACCTACTCTCAAT 5406
 Db |||||
 QY 5147 TCTTTGATATAAAGCTGATTAATTTTGTATTAATTTGAATGAAGACCTACTCTCAAT 5206
 Db |||||
 QY 5407 AGATAAATCTTTTGGATTATATCAAGAAATTTTGCCTTTTCTTCTCAAAATTAATPACA 5466
 Db |||||
 QY 5207 ACATACAAAGCTTTT-GATTAATTTAGAAATTTTGCCTTTTATTTCTTCTCAAAATTTCA 5265
 Db |||||
 QY 5467 TATGTATGATATATATCCATATATAGTTTTCCTGATTAATGATTAATTAATTAATTA 5526
 Db |||||
 QY 5266 TATCTATATA-TATATATACACATATAATATATATATTTTCTG---ATTAAATATA 5321
 Db |||||
 QY 5527 TTGCGGTGCTTCCAGACTTTTGTCTATATTTAAGTATATTTTATATAGCAAGAA 5586
 Db |||||
 QY 5322 TTGCGGTGCTTTAGGACTTTATGCTTCTCTATTTAAGTCTGTTGTTTCCATAGCAACAT 5381
 Db |||||
 QY 5587 CATATCTGAA--TGTTTATAATCTTTAATATTTATATATAGTAAATTTTGTG 5641
 Db |||||
 QY 5382 ATGATTTGGAATGTGTTTATAAATCTTTGCTAATATAATATAATAGATAATTTTGTG 5438
 Db |||||

RESULT 8

BC042669

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BC042669 5480 bp mRNA linear ROD 07-OCT-2003
 Mus musculus RIKEN cDNA B130024H06 gene, mRNA (cDNA clone MGC:28649
 IMAGE:4235024), complete cds.

BC042669
 MGC
 Mus musculus (house mouse)

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 5480)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

FEATURES

source

1..5480
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="MGC:28649 IMAGE:4235024"
 /tissue_type="Kidney, normal. 5 month old male mouse."
 /clone_lib="NCI CGAP_Kid14"
 /lab_host="DH103"
 /note="Vector: pCMV-SPORT6"
 1..5480
 /gene="B130024H06Bik"
 /note="synonym: B130031115"
 /db_xref="LocusID:218441"
 /db_xref="MGI:2145181"
 150..4736
 /codon_start=1
 /product="RIKEN cDNA B130024H06"
 /protein_id="RAH42669.1"
 /db_xref="GI:27503688"
 /translation="MDSYFKAASVGLDKLLDDFQNPDKQYLDPAFYAFNOCSSVSS
 ASPAQLIKQRCISTCASACEDANETFLGKHEGLTFRFNKNVAGLDLIS
 VDASTSDIQSPKRCSPKPCVDITSDMNLVHATNSBEDIKQLLPDPKSSADTLIT
 LSSSVSEALTIVSSVDCGNAVBQNNINAGIKNRDITSIKELGVKVDMAFDSCCKN

gene

CDS

Hopkins, R.P., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Schetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carrincci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, J., Smillius, D.E.,
 Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 2238257
 12477932
 2 (bases 1 to 5480)
 Strausberg, R.
 Direct Submission
 Submitted (02-JAN-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs@mail.nih.gov
 Tissue procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louised, H.,
 Kowis, C.R., Speed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 37 Row: k Column: 18
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 27734995.

RTNLDKDIISLETVDFMPSVLMQSEMSNTKDNPOYKRLPCELLKDDGCLAE
 KVAVANNTECLEGGSSNTIAMPCKLPENEGISPDSPASKOENFXPLPDPLOBRNRS
 VFNKQTVKSDNRDLKNDNDIIVHSGDDVPPSLSCLSLGSGLCSLHNEHSDIILP
 PNESEGONDAVTHIEIOKSDVLDEGTDLSSKETCRSIFLOPVNEKKEGVEVEEM
 VIGSELSEPEDASSAAGSPVALSAASVPEAPGPCCEGLTPSSDMDCOEIDYFNID
 ESMRSGLISDAELDAFLKEOCLSNSTMSAGENVDSOLOKNOIIMTKGLHDEAGDI
 YFNAAGAENGCGVNCSTSDXENTENNGLSIGEKAIPTRELSACOPDIRDELVP
 PSIKTAGVAGRPKQLSLUPPGPRSRSKLNKLPDVPDPSEPCATANATAVSCSADHI
 PDSQVFNISYIDIESNFEDGSSFTVANKDSLPENRKESVLGQKOPTWPDSEAPN
 CMNCQVFTFKRHHRCACGKVFQVCCNRKQLYLEKEARVICVETINKQAQF
 ERNMSGSLCKRNHNECATDQLOBTOTSTPSPTLPIGALKQVNEVGPSCKEQK
 RVFADGLPENGVAOTTKLSSGKRCSDPSVLPDVTMINKVDRTHSPVKEPNN
 GLGDIIRSELQSPTCHTAPVESLPGNTGTEGLMPGPBFTLSDDDVDEEESTPTVV
 PANSGLPVASISDIRLCSVAKVCVNNISLLPDDDIGLPLLATSEDSQVFWQERP
 SHEQIILLSEGGEPFATFVLNANLVNVLKLYISSEKYWFSTNHLGHLGQABLIIV
 LOCLPNEDVPKDIIFLFIITYDALKGYIENLDNLTTFESFNSKHGGFLFIPT
 FQNLGILPVSPFLCIGLILKLEIPWAKVFNMLRLGAEYKAPALPTSVRGRKP
 LFEIGHTIMLLVDRNYOYTLHNDIQLLHMKGSKIKIPRKYSVDMVKVHSNN
 EHVISGASPTSEADSHLNCQSDGVYTOANSATGQPRKVTGASVFNENGALKTSSG
 FLAKSSIVEDGLVQITPETMEGLRLALREQDFRIQCKVDAVDLREXVDLCWVDS
 ERNKGVSIGVSSVEGFESEKIKUETDETEEKIVKCTEVEYFPLKDDIISLSSY
 QFAKEIYAVCSAALCPHLRLTKSNRMKIGLRVSDTDMVFOAGCEGQLLPOHYLND
 LDSALIPVHIGGTSNSLPLEIEIAFFILENSE"

misc_feature

2337..2531

/note="FYVE; Region: FYVE zinc finger. The FYVE zinc
 finger is named after four proteins that it has been found
 in: Fab1, YORF/ZK632.12, Vac1, and BEAL. The FYVE finger
 has been shown to bind two Zn++ ions. The FYVE finger has
 eight potential zinc coordinating cysteine positions. Many
 members of this family also include two histidines in a
 motif R+HHC+XGG, where + represents a charged residue and
 X any residue. We have included members which do not
 conserve these histidine residues but are clearly related"
 /db_xref="CDD:pfam01363"

ORIGIN

Query Match 50.3%; Score 3333; DB 10; Length 5480;
 Best Local Similarity 78.8%; Pred. No. 0;
 Matches 4298; Conservative 0; Mismatches 1020; Indels 133; Gaps 22;
 Qy 208 AGGCATACAGAAATAAATCTGTAATAGTCTGCAGTAGATGGACAGTATTTTAAAG 267
 Db 109 AGACATAGAAATAGTGTCTAGTAGAGCCACAGGAGGATGGACAGTACTTTTAAAG 168
 Qy 268 CAGCTGTCACTGACCTCGACAACTCTCTGATGATTTTGAACAGACCAGATGAACAAG 327
 Db 169 CAGCTGTCACTGAGCTTGAACAACTCTCTGATGATTTTGAACAGATCCAGATAAGCAAG 228
 Qy 328 ATTAATCTCGAGATGTACAAAATGCATATGATTTCAACACTGCTCAGTTTCTTCAGAT 387
 Db 229 ATTAATCTCCAG-----ATGCATATGCTTTTAAACCACTGCTCAGTTTCTTCAGAT 279
 Qy 388 TGGCTTCTCACAGGAACTTCATGCTCCCAAGACCAAGAGTGGCTTAATAGTTGTG 447
 Db 280 CGGCCTCTCCACACTAGCT---TTACTGTCAAAAGGACCAAGATGATCAGTACTTTGTG 336
 Qy 448 CCTCATCAGAAACCAAGTATGGA---ACAAATGAGAGTTCCTCGAATGAAACCACTCA 504
 Db 337 CCTCATCAGAGCATGCTGTGAAGATGCAATGAGACTTTTCTTGAAGGAAGATACACG 396
 Qy 505 AGGCACTTACTTCTATACAAAATGAAAATATGTAACAGACTTGATCTTCTTCTCTG 564
 Db 397 AGGCCTTAACCTCTAGACCGAATGAAAACACGTCAGCGGACTTGATCTTCTTCTCTG 456
 Qy 565 TGGATGGTGGTACTTTCAGATGAAATCCAGCCGTTATATATGGGACGATGTAGTAAACCTA 624
 Db 457 TGGATGGAGTACTTTCAGATGAGATCCAGCTTCTCTGTATGAGACGCTGTAGTAAGCTG 516
 Qy 625 TCTGTGATCTGATAGTACATGGGTAACTTAGTTCATGCAACCAATAGTGAAGAGATA 684
 Db 517 TCTGTGACCTGATAAGTGACATGGGTAACTTAGTTCATGCAACCAATAGTGAAGAGACA 576
 Qy 685 TTAAAAAATTATTCAGAGATGATTTTAAAGTCTAATGAGATTCCTCTGATTGGATTGGATT 744

Db 577 TTAAGCAGTATATGCGCAGATGACCCCTAAGTCTTCTGCAGATACTTTGATCACTTGGACT 636
 Qy 745 TATCTTTCAGTGTGCAGATACTCCCTGTGTGTCTTCTCAACAGACCATGATAGTATGTGCA 804
 Db 637 CGTCTTCAGTTTCAGAGGCTCTCACTGTCTCTTTCAGTAGACTGTGGTAGTAATGTGCTGCA 696
 Qy 805 GAGAACACAGAGATGATATCAGTCTGAAATACAAAAATAGAGAAATCGAGGAATCAAG 864
 Db 697 GAGAGAGACAGAAATACATCAACCGCTGTGTATAAAAAAGCAGAGATAT---TAGTATCAAG 753
 Qy 865 AATTCGGGTATAAAAAGTAGATACAAACACTTTCAGATTCCTATAATTACAGTGGAAACAGAAA 924
 Db 754 AATTCGGGTATAAAAAGTAGATGACATGCGACATTTTGAATCTTCTGTAATATAATAGACAGAAA 813
 Qy 925 AATTTAAAGATAAAAAGATCTTTTAATCAGTGTAGATCAATTTGTGATTTTAACTATGTCAT 984
 Db 814 AATTTAAAGATAAAAATTTATTTCTAATGATGTAGAAACA---GTTGATTTTTCAGATGCCAT 870
 Qy 985 CTGCTTTGACTCGACAAAGTTTCCAAAATGTTTCATGTCAAAAGACAAAGCTACAAACAAGA 1044
 Db 871 CTGTTTGTATGGAACAAAGTTTCTGAAATGTCTAATACCAAGATTAACCCACATACAAGA 930
 Qy 1045 GCCAGCCATGTGGATTACTAAAAAGATGTGGCTTAGTAAAAAGAGAGTAGATGGCAG 1104
 Db 931 GGCTGCCATGTCAATTTGTTTAAAGATGATGGCTGTTTGCAGAGAGAAAGTAGCTGTGG 990
 Qy 1105 TCATAACTGCCCGCAGAAATGTTTAAAGAGAGAGG-----GCAAGACAGTGTCTTGAACCT 1158
 Db 991 CAGTCAATAATACAGAGTGTTTTGAAGAGAGAGGAGGAGTAACAAATTTGCTATGCTTCT 1050
 Qy 1159 GCAGCTTCCGAAAAAATGAAGATTTTATGCTTAAATGATTCAAAATTCAGAGATGAAAAAT 1218
 Db 1051 GCAAGCTTCCAGAAATGAAGGGATATCCCAAGTGCACCCAGCTTCAAAAGATGAAAAAT 1110
 Qy 1219 TCAATATCACTGACTTTTCCCTTTTCAGGAAGATAAGACT-----GTTATAAACATCTG 1272
 Db 1111 TCAAGTTACTGACTTTTCCCTTCCAGGAAAAATAGAACTTCTGTGTCTATGAACAGACTG 1170
 Qy 1273 CACAAGAGACTCAAAAAAGTTTAGACCTTAAAGGATAATGATGATAATCAAGATTCCTCTT 1332
 Db 1171 TAAAGAGACTCAAGAAATTTAGACCTTAAAGATAATTAATGATAT----- 1216
 Qy 1333 CAGCTTTACATGTTTCCAGTAAAGATGTCGGCTCTCATTTGTCCTGTCTTCTCGGTCTG 1392
 Db 1217 ---AGTCCATGTTTCAGGTGATGATGTGCCCTTCACTGTCTGTCTTTTCAATGTCTG 1272
 Qy 1393 GGTCTATGTGTGGATCATTAATTTGAAGTAAAGACGCGGGTGTATTTTACCTCAGCATG 1452
 Db 1273 GGTCTTTGTGTGGATCGTTTAAATTCATAATATGAACACAGTGTATTTTACCTCCTAATG 1332
 Qy 1453 AACATAAAGATTAATATACAAGATGCAGTGAATATACATGAAGAAATACAGAAACAGTTG 1512
 Db 1333 AGTCTGAAGGCCAGAACAAATGATGACGTGACTATACATGAAGAGATACAGAAGATGATG 1392
 Qy 1513 TTTCTAGTGGGAAACCAATTCAAAAGAGATGATCTTTTGAACAGGAAAAATGTAAAGCA 1572
 Db 1393 TTTTAGATGGG-----GAGACAGATCTTTTCAAGAGGAAACGCTGTAGAGCA 1440
 Qy 1573 TACTCCTTCAGTCAATTAATTTGAAGGATGGAAGACAGAAAGATAGATCCTGACCAAGACAG 1632
 Db 1441 TATTTCTTCAGCCAGTAATTAATGAAGAGAGGGAGGAAAGGTTGAGGTTGAGGAGATG 1500
 Qy 1633 TAATCAGAGCTGAGTCTTTTGGG-----TGGTGGTGAACCAAGTCTTACAGTTGTAGAAAT 1686
 Db 1501 TAATCAGTGGTGAATCTTTTGGAGTCCCTTGAGAGCGCCAGCTCTGCTGCGAGCAGCTGGGT 1560
 Qy 1687 CTCAGAGGGGCTTTCTGSCACTCATGTCAGAGTCTTCTGATTTGTTGTTGAGGTTTGA 1746
 Db 1561 CTCGGTGGCACTTTCTGCTGCAAGTGTCCCAAGAGCTCTGCTGCTCATGTGAAGGTCT--- 1618
 Qy 1747 TTAATACTTTTCAAGCAATGATATGGATGGCAAGACTTAGATTAATTAATGATG 1806

/tissue type="human retina"									
/clone_lib="686 (synonym: hlcc3). Vector pSport1_sfi; host									
DH10B; sites SfiIA + SfiIB"									
/dev stages="adult"									
5117..5122									
polyA_signal									
polyA_site									
5140									
ORIGIN									
Query Match									
Best Local Similarity 40.0%; Score 2655.2; DB 9; Length 5158;									
Matches 2668; Conservative 0; Mismatches 3; Indels 1; Gaps 1;									
QY	3939	GACCTTCGAATATACCACTGATACCTTGATATATAGATCAATCACTGTTGATTCATATGGA	3998						
DB	2473	GACCTTCGAATATACCACTGATACCTTGATATATAGATCAATCACTGTTGATTCATATGGA	2532						
QY	3999	ATGGGAAAAGCTGCATAAAAATACCGGAAAAGTACAGTGATGTAATGAAAGTACTA	4058						
DB	2533	ATGGGAAAAGCTGCATAAAAATACCGGAAAAGTACAGTGATGTAATGAAAGTACTA	2592						
QY	4059	AATCTTCCATGAGCATGTCATTAGCATTTGAGCAAGTTTCAGTACAGAGCATGCT	4118						
DB	2593	AATCTTCCATGAGCATGTCATTAGCATTTGAGCAAGTTTCAGTACAGAGCATGCT	2652						
QY	4119	CATCTAGTCTGTATACAGAAATGATGGAAATTTATGAAACACAGGCCAACAGTCCACTGGC	4178						
DB	2653	CATCTAGTCTGTATACAGAAATGATGGAAATTTATGAAACACAGGCCAACAGTCCACTGGC	2712						
QY	4179	CATCTAGAAAAGTACAGGTGCAAGTTTGTGTATTCATGAGCTCTAAAAACATCT	4238						
DB	2713	CATCTAGAAAAGTACAGGTGCAAGTTTGTGTATTCATGAGCTCTAAAAACATCT	2772						
QY	4239	TCAGATTTCTTGCTAGCTGACGATAGTTCAGATGCTTAATGTTACAAATAACTCCA	4298						
DB	2773	TCAGATTTCTTGCTAGCTGACGATAGTTCAGATGCTTAATGTTACAAATAACTCCA	2832						
QY	4299	GAGACCATGAATGGCTTCGCGCTAGCTTTACGAGAACAGAAAGTCTTAAAAATPACATGT	4358						
DB	2833	GAGACCATGAATGGCTTCGCGCTAGCTTTACGAGAACAGAAAGTCTTAAAAATPACATGT	2892						
QY	4359	GGGAAAGTTGATGCTAGCTGACGATGCTGAGATACCTGGATCTCTGCTGGTAGATGCTGAA	4418						
DB	2893	GGGAAAGTTGATGCTAGCTGACGATGCTGAGATACCTGGATCTCTGCTGGTAGATGCTGAA	2952						
QY	4419	GAAAAGGAAAACAAAGGAGTTATCAGTTTCAGTGGATGGAATATCATTTACAGGATTTCCA	4478						
DB	2953	GAAAAGGAAAACAAAGGAGTTATCAGTTTCAGTGGATGGAATATCATTTACAGGATTTCCA	3012						
QY	4479	AGTGAAAAAATAAATCTGGAAGCAGATTTTGAACCCGATGGAAGATGTAATAATGTACC	4538						
DB	3013	AGTGAAAAAATAAATCTGGAAGCAGATTTTGAACCCGATGGAAGATGTAATAATGTACC	3072						
QY	4539	GAGGTGTTCTACTTTCTTAAAGGACAGGATTTATCTATTTTATCAACTTCTTATCAGTTT	4598						
DB	3073	GAGGTGTTCTACTTTCTTAAAGGACAGGATTTATCTATTTTATCAACTTCTTATCAGTTT	3132						
QY	4599	GCAAAAGAAAATAGCCATGGCTTTGATGCTGGCTGTGCTCCCTACCTGAAAACCTCTAAAA	4658						
DB	3133	GCAAAAGAAAATAGCCATGGCTTTGATGCTGGCTGTGCTCCCTACCTGAAAACCTCTAAAA	3192						
QY	4659	AGTAATGGATGAATAAATTTGGACTCAGAGTTTCCATTTGACACTGATATGTTGAATTT	4718						
DB	3193	AGTAATGGATGAATAAATTTGGACTCAGAGTTTCCATTTGACACTGATATGTTGAATTT	3252						
QY	4719	CAGCAGGATCTGAGGCCAACTTCTGCCTCAGCATTTATCTTAAATGATCTTGTAGTGCT	4778						
DB	3253	CAGCAGGATCTGAGGCCAACTTCTGCCTCAGCATTTATCTTAAATGATCTTGTAGTGCT	3312						
QY	4779	CTGATACCTGTGATCCATGGTGGGACCTCCAACTCTAGTTTACCATTAGAAAATGAAATTA	4838						
DB	3313	CTGATACCTGTGATCCATGGTGGGACCTCCAACTCTAGTTTACCATTAGAAAATGAAATTA	3372						
QY	4839	GTGTTTTTCATTATAGAACATCTTTTTTAGTGAAGAAATGTGCCATTTACATATTGCAA	4898						

DB	3373	GIGTTTTTCATTATAGACATCTTTTTTAGTGAAGAAATGTGCCATATTACATATTGCAA	3432						
QY	4899	CCTAAATTTGTTAAAACTAATCCACACATAAGCTGAAAGCTGAAATGCCACAAACACTAAAAGTAT	4958						
DB	3433	CCTAAATTTGTTAAAACTAATCCACACATAAGCTGAAAGCTGAAATGCCACAAACACTAAAAGTAT	3492						
QY	4959	AAATATGCTGATTTTGAACACACATAAGCTTTGCTCTTTTAGCGAGGAATGATCTTTTCA	5018						
DB	3493	AAATATGCTGATTTTGAACACACATAAGCTTTGCTCTTTTAGCGAGGAATGATCTTTTCA	3552						
QY	5019	AATCATTTAGCACAAATATTAAATATCTAAAAATTTAAGAGATCCATATCTTCTGTAGCTT	5078						
DB	3553	AATCATTTAGCACAAATATTAAATATCTAAAAATTTAAGAGATCCATATCTTCTGTAGCTT	3612						
QY	5079	TACAATTAATTTAAGTACTTAAAGACACAGGATTTCTTTTAAAGAAATTTATAGCATTTAC	5138						
DB	3613	TACAATTAATTTAAGTACTTAAAGACACAGGATTTCTTTTAAAGAAATTTATAGCATTTAC	3672						
QY	5139	TGTGTTATTTAAATGCTAAGCCAAAGTATCTGCACTTAGGTATACCTCTTTATGCCAATA	5198						
DB	3673	TGTGTTATTTAAATGCTAAGCCAAAGTATCTGCACTTAGGTATACCTCTTTATGCCAATA	3732						
QY	5199	ATGATTTTAAATGAAGCTCTTTTCAAGATGTAACCTTTAAGGAAATATCTGCTTTGTGT	5258						
DB	3733	ATGATTTTAAATGAAGCTCTTTTCAAGATGTAACCTTTAAGGAAATATCTGCTTTGTGT	3792						
QY	5259	ATATGCCAGTTAGATATCTGTTTCTTAAAGCTGTCAAATTTGATTTTCAGTGGGACAAAA	5318						
DB	3793	ATATGCCAGTTAGATATCTGTTTCTTAAAGCTGTCAAATTTGATTTTCAGTGGGACAAAA	3852						
QY	5319	ACCAGTTTGGAGCTCTTAGACTTATAATCTTTTGAATAAAACCTGATACTTTATTGTATA	5378						
DB	3853	ACCAGTTTGGAGCTCTTAGACTTATAATCTTTTGAATAAAACCTGATACTTTATTGTATA	3912						
QY	5379	ATTGGAGTGGAGACCTTACCTCCATAATTTAGATAAACTCTTTTGGATTATATATCAGAAAT	5438						
DB	3913	ATTGGAGTGGAGACCTTACCTCCATAATTTAGATAAACTCTTTTGGATTATATATCAGAAAT	3972						
QY	5439	TTGCCCTTTTTCTCTCATAATTTATACATATGATGTTATATATATCCACATATATAGTT	5498						
DB	3973	TTGCCCTTTTTCTCTCATAATTTATACATATGATGTTATATATATCCACATATATAGTT	4032						
QY	5499	TTCCCTGTATTAATGGATTTAAATAATTTGCGGTGCTTCAGGACTTTTTCCTCTATA	5558						
DB	4033	TTCCCTGTATTAATGGATTTAAATAATTTGCGGTGCTTCAGGACTTTTTCCTCTATA	4092						
QY	5559	TTTAAAGTATTTGTTTATAGCAAGAACATATCTGAATGTTTATATAAACTTTTATAAA	5618						
DB	4093	TTTAAAGTATTTGTTTATAGCAAGAACATATCTGAATGTTTATATAAACTTTTATAAA	4152						
QY	5619	TTTATATGTAAGTAAATTTTTTGTATCAATGCAATATTTTTTTCCTCCTTCTTCC	5678						
DB	4153	TTTATATGTAAGTAAATTTTTTGTATCAATGCAATATTTTTTTCCTCCTTCTTCC	4211						
QY	5679	AAACATATACCTGTATTTTACCACTTCTAAGAGTCACTGACGACGGGCGCAGATGACCTT	5738						
DB	4212	AAACATATACCTGTATTTTACCACTTCTAAGAGTCACTGACGACGGGCGCAGATGACCTT	4271						
QY	5739	GAGTAGTCAATTTATGCAATTAATGAAGCCTGAAAACAGGTTTTTTTACTTCCACTTA	5798						
DB	4272	GAGTAGTCAATTTATGCAATTAATGAAGCCTGAAAACAGGTTTTTTTACTTCCACTTA	4331						
QY	5799	ATCCCTTAGAAATTTCTTGGCACTTCCGATATTTTTCATTGACCTGGTGTATATAGTATAA	5858						
DB	4332	ATCCCTTAGAAATTTCTTGGCACTTCCGATATTTTTCATTGACCTGGTGTATATAGTATAA	4391						
QY	5859	ATTTAAATGAACTAATTTACTTTTGCATATTTTAAATTTCTTTATATGTTAGTTATTTTAA	5918						
DB	4392	ATTTAAATGAACTAATTTACTTTTGCATATTTTAAATTTCTTTATATGTTAGTTATTTTAA	4451						
QY	5919	TAAACAGGATTTTAAACATAAGTTTAAATCTTGTATTTGAAATTTGTTACAGAGCTTCTC	5978						

in: Fabi, YOTB/ZK632.12, Vac1, and BEA1. The FIVE finger has been shown to bind two Zn++ ions. The FIVE finger has eight potential zinc coordinating cysteine positions. Many members of this family also include two histidines in a motif R-HHC+XCG, where + represents a charged residue and X any residue. We have included members which do not conserve these histidine residues but are clearly related" /db_xref="CDD:pfam01363"

ORIGIN

Query Match 37.0%; Score 2454.4; DB 9; Length 2902;
 Best Local Similarity 99.2%; Pred. No. 0;
 Matches 2467; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY	185	ATATCTCTCCAGGATGTTCTCAAGGCATACAGAAATTAATTCCTGAATAGTCTGCAGG	244
Db	47	AAAGCTCCGAGGGCTGTAGGAGGCNTACAGAAATTAATTCCTGAATAGTCTGCAGG	106
QY	245	TAGATGACAGATTAATTTAAAGCAGTGTCTGAGTGTGGACAAATCTCTTGATGATTT	304
Db	107	TAGATGACAGATTAATTTAAAGCAGTGTCTGAGTGTGGACAAATCTCTTGATGATTT	156
QY	305	TGAACAGAACCCAGATGACAGAGATTATCTCGAGATGTACAAATGCATATGATCTAA	364
Db	167	TGAACAGAACCCAGATGACAGAGATTATCTCAAGATGTACAAATGCATATGATCTAA	226
QY	365	CCACTGCTCAGTTTCTTCAGAGTTGGCTTCTCCACAGCACTTCATTGTCTCCAAAAGA	424
Db	227	CCACTGCTCAGTTTCTTCAGAGTTGGCTTCTCCACAGCACTTCATTGTCTCCAAAAGA	286
QY	425	CAAGAGTGGTTAATGTTGCTCTATCGAAGCAAGCTATGGAACAAATGAGGTTTC	484
Db	287	CAAGAGTGGTTAATGTTGCTCTATCGAAGCAAGCTATGGAACAAATGAGGTTTC	346
QY	485	CCTGAATCAAAAACACCTCAAGGACCTTACTTCTATACAAATCAAAAATGTAAACAGG	544
Db	347	CCTGAATCAAAAACACCTCAAGGACCTTACTTCTATACAAATCAAAAATGTAAACAGG	406
QY	545	ACTTGATCTTCTTCTCTGAGTGGTGGTCTATCAGATGAATCCAGCGGTTATAT	604
Db	407	ACTTGATCTTCTTCTCTGAGTGGTGGTCTATCAGATGAATCCAGCGGTTATAT	466
QY	605	GGGAGATGTAGTAACCTATCTGTGATCTGATAGTGACATGGTAACTTAGTTCATGC	664
Db	467	GGGAGATGTAGTAACCTATCTGTGATCTGATAGTGACATGGTAACTTAGTTCATGC	526
QY	665	AACCAATAGTGAAGAAGATATTAATAAATTTATCCAGATGATTTTAAGTCTAATGCAGA	724
Db	527	AACCAATAGTGAAGAAGATATTAATAAATTTATCCAGATGATTTTAAGTCTAATGCAGA	586
QY	725	TTCTTGATGGATGGATTTATCTTCAGTGTGAGATCTCCCTGTCTTCTTCAACAGA	784
Db	587	TTCTTGATGGATGGATTTATCTTCAGTGTGAGATCTCCCTGTCTTCTTCAACAGA	646
QY	785	CCATGATAGTATCTGTGCAGAACCAACAGAAATGATATCAGTTCTCGAATTAACAAAATAG	844
Db	647	CCATGATAGTATCTGTGCAGAACCAACAGAAATGATATCAGTTCTCGAATTAACAAAATAG	706
QY	845	AGAAATCGGAGGATCAAGAAATTTGGGTATATAAAGTAGATACACACTTTCAGATTCCTTA	904
Db	707	AGAAATCGGAGGATCAAGAAATTTGGGTATATAAAGTAGATACACACTTTCAGATTCCTTA	766
QY	905	TAAATTCAGTGCACAGAAAATTTAAAGATATAAAGATCTTTAATCAGTTAGAAATCAAT	964
Db	767	TAAATTCAGTGCACAGAAAATTTAAAGATATAAAGATCTTTAATCAGTTAGAAATCAAT	826
QY	965	TGTTGATTTTAACATGTCATCTGTTGATCTGACGACAAAGTTTCCAAAATGTTTCATGCCAA	1024
Db	827	TGTTGATTTTAACATGTCATCTGTTGATCTGACGACAAAGTTTCCAAAATGTTTCATGCCAA	886
QY	1025	AGACAGCTTACACACAGAGCCAGCCATGTGGATTACTAAAGATGTGGCTTAGTAAA	1084
Db	887	AGACAGCTTACACACAGAGCCAGCCATGTGGATTACTAAAGATGTGGCTTAGTAAA	946

QY	1085	AGAGGAAGTAGATGTGGCAGTCATAACTGCCGAGAAATGTTTAAAGAAAGAGGGCAAGAC	1144
Db	947	AGAGGAAGTAGATGTGGCAGTCATAACTGCCGAGAAATGTTTAAAGAAAGAGGGCAAGAC	1006
QY	1145	AAAGTGTCTTGACCTCGACGCTTCCGAAAATTAAGAAATTTATGCTTAAATGATCTAAATTC	1204
Db	1007	AAAGTGTCTTGACCTCGACGCTTCCGAAAATTAAGAAATTTATGCTTAAATGATCTAAATTC	1066
QY	1205	AAAGATGAATAATTTCAATTAACCTGACTTTTCTTTTCAGGAAGATAAGACTGTTTATAAA	1264
Db	1067	AAAGATGAATAATTTCAATTAACCTGACTTTTCTTTTCAGGAAGATAAGACTGTTTATAAA	1126
QY	1265	ACAATCTGCACAAGAAGACTCAAAAAGTTTACACCTTAAAGGATTAATGATTAATCCAAAGA	1324
Db	1127	ACAATCTGCACAAGAAGACTCAAAAAGTTTACACCTTAAAGGATTAATGATTAATCCAAAGA	1186
QY	1325	TTTCTCTTCAGCTTTTACATGTTTCCAGTAAAGATGTGGCTCTCATTTGCTCTCTCTTCC	1384
Db	1187	TTTCTCTTCAGCTTTTACATGTTTCCAGTAAAGATGTGGCTCTCATTTGCTCTCTCTTCC	1246
QY	1385	TGCGTCTGGGTCTATGTGTGGATCAATTAATTGAAGTAAAGCACGGGTGATTTTTTACC	1444
Db	1247	TGCGTCTGGGTCTATGTGTGGATCAATTAATTGAAGTAAAGCACGGGTGATTTTTTACC	1306
QY	1445	TCAGCATGAACAATAAGATAATATACAGATGTCAGTGACTATACATGAAGAAATACAGAA	1504
Db	1307	TCAGCATGAACAATAAGATAATATACAGATGTCAGTGACTATACATGAAGAAATACAGAA	1366
QY	1505	CAGTGTGTTCTTAGTGGGGAACCAATCAAGAGAGATGATCTTTTGAACAGGAAAAATG	1564
Db	1367	CAGTGTGTTCTTAGTGGGGAACCAATCAAGAGAGATGATCTTTTGAACAGGAAAAATG	1426
QY	1565	TAAGAAGCATACTCTTCAGTCAATTAATTGAAGGATGAAGACAGAAAGATAGATCTTGA	1624
Db	1427	TAAGAAGCATACTCTTCAGTCAATTAATTGAAGGATGAAGACAGAAAGATAGATCTTGA	1486
QY	1625	CCAGACAGTAAATCAGAGCTGAGTCTTGGATGGTGGTGACACCCAGTCTACAGTTGTAGA	1684
Db	1487	CCAGACAGTAAATCAGAGCTGAGTCTTGGATGGTGGTGACACCCAGTCTACAGTTGTAGA	1546
QY	1685	ATCTCAAGAGGGGCTTCTGCGACTCATGTCCAGAGTCTTCTGATGTTGTGAAGGTTT	1744
Db	1547	ATCTCAAGAGGGGCTTCTGCGACTCATGTCCAGAGTCTTCTGATGTTGTGAAGGTTT	1606
QY	1745	TATTAATACTTTTCAAGCAATGATATGGATGGGCAAGACTTAGATTAATTAATTTGA	1804
Db	1607	TATTAATACTTTTCAAGCAATGATATGGATGGGCAAGACTTAGATTAATTAATTTGA	1666
QY	1805	TGAAGCGCAAAAAGTGGCCCACTAATTAATGATGATGCTGAACCTTGAATGCTTTCGACAGA	1864
Db	1667	TGAAGCGCAAAAAGTGGCCCACTAATTAATGATGATGCTGAACCTTGAATGCTTTCGACAGA	1726
QY	1865	ACAGTATCTTTCAGACCACTAACAATAAAGTCTTTTGAAGAAATGTAAATGATCTTAAATC	1924
Db	1727	ACAGTATCTTTCAGACCACTAACAATAAAGTCTTTTGAAGAAATGTAAATGATCTTAAATC	1786
QY	1925	GCAATGATATCAGATAGATATGAAAAGGCTTAGATGATGAAAACATCAATAATATATTT	1984
Db	1787	GCAATGATATCAGATAGATATGAAAAGGCTTAGATGATGAAAACATCAATAATATATTT	1846
QY	1985	CAATGCAGAGCAGAGCTTATGGGAAAGTCAATGGTATTAATATAATTTGTGAAACAGT	2044
Db	1847	CAATGCAGAGCAGAGCTTATGGGAAAGTCAATGGTATTAATATAATTTGTGAAACAGT	1906
QY	2045	TGATTAACAAAATCAATAGAAAATGGCTTTCTTTAGGAGAAAACACACTATTCCAGT	2104
Db	1907	TGATTAACAAAATCAATAGAAAATGGCTTTCTTTAGGAGAAAACACACTATTCCAGT	1966
QY	2105	TCAACAAGGGTTTACCTACCGAGTAACTCTGAGATTAACAAATCAATTTATCAGTCTCTGATAT	2164
Db	1967	TCAACAAGGGTTTACCTACCGAGTAACTCTGAGATTAACAAATCAATTTATCAGTCTCTGATAT	2026

```
QY 2165 TAAAGTCAATCTGTTGGAGGGGCGAGACCTAAGCAATTTGTTAGCTTCCATCAAGAAC 2224
Db 2027 TAACAGTCAATCTGTTGGAGGGGCGAGACCTAAGCAATTTGTTAGCTTCCATCAAGAAC 2086
QY 2225 AAGAGTTCAGAGACCTGAATTAAGCAGATGTTCCAGATACAAATAGAAAGTGAACCCAG 2284
Db 2087 AAGAGTTCAGAGACCTGAATTAAGCAGATGTTCCAGATACAAATAGAAAGTGAACCCAG 2146
QY 2285 CACAGCAGATACCGTTGTTCCCAATCACTTGCTGTATAGATTCTACAGCTGATCCACAGGT 2344
Db 2147 CACAGCAGATACCGTTGTTCCCAATCACTTGCTGTATAGATTCTACAGCTGATCCACAGGT 2206
QY 2345 TAGCTCAACTCAATTAATACATTAATGATATAGAAAGTAATTTGAAAGGTGGATCTAGTTTGGT 2404
Db 2207 TAGCTCAACTCAATTAATACATTAATGATATAGAAAGTAATTTGAAAGGTGGATCTAGTTTGGT 2266
QY 2405 AACTGCAAAATGAGATTCTGTAACCTGAACACCTTGCAAGAGAGGCTTGTTTGGGCCA 2464
Db 2267 AACTGCAAAATGAGATTCTGTAACCTGAACACCTTGCAAGAGAGGCTTGTTTGGGCCA 2326
QY 2465 GAAACAGCTTACTGGGTTCTCTGATTGAGAGCTCCAACTGTATGAATGCCAAGTCAA 2524
Db 2327 GAAACAGCTTACTGGGTTCTCTGATTGAGAGCTCCAACTGTATGAATGCCAAGTCAA 2386
QY 2525 ATTACTTTTACCAAGCGGCGACACCACTTCCGAGCATGTCGGAAGATTTTGTGGTGT 2584
Db 2387 ATTACTTTTACCAAGCGGCGACACCACTTCCGAGCATGTCGGAAGATTTTGTGGTGT 2446
QY 2585 CTGTTGTAATAGGAAGTGAATCTGCAATCTAGAAAGAGCAAGAGATGATGCTAGT 2644
Db 2447 CTGTTGTAATAGGAAGTGAATCTGCAATCTAGAAAGAGCAAGAGATGATGCTAGT 2506
QY 2645 CTGCTATGAACACTTATAGTAAAGCTCAG 2672
Db 2507 CTGCTATGAACACTTATAGTAAAGCTGAG 2534
```

RESULT 11

```
AC008382/c
LOCUS 252015 bp DNA linear PRI 23-OCT-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-218P2, complete sequence.
ACCESSION AC008382
VERSION AC008382.6 GI:16328245
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 252015)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 252015)
DOE Joint Genome Institute.
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 252015)
DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Oct 23, 2001 this sequence version replaced gi:9255963.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.6% of Sequence;
Estimated Total Number of Errors is 0.7.
Location/Qualifiers
FEATURES
source 1..252015
/mol_type="genomic DNA"
```

```
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-218P2"
```

ORIGIN

```
Query Match 34.0%; Score 2254.4; DB 9; Length 252015;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2258; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 316 CAGATTGAACAAGATTATCTCGCAGATGTACAAAATGCAATATGATTTCTAACCACTCTCTCAG 375
Db 20953 CAGATTGAACAAGATTATCTCCAAGATGTACAAAATGCAATATGATTTCTAACCACTCTCTCAG 20894
QY 376 TTTCTTCAGAGTTGGCTTCTTCACAGCGAATCTTCTCTCTCCAAAAGACCAAGAGTGGG 435
Db 20893 TTTCTTCAGAGTTGGCTTCTTCACAGCGAATCTTCTCTCTCCAAAAGACCAAGAGTGGG 20834
QY 436 TTAATAGTTGTGCTTCATCAGAAACAAGCTATGGAACAAATGAGAGTTCCCTGAAATGAA 495
Db 20833 TTAATAGTTGTGCTTCATCAGAAACAAGCTATGGAACAAATGAGAGTTCCCTGAAATGAA 20774
QY 496 AAACACTCAAGGAGCTTACTTCTATACAAAATGAAAAAAATGTAAACAGAGCTTGATCTTC 555
Db 20773 AAACACTCAAGGAGCTTACTTCTATACAAAATGAAAAAAATGTAAACAGAGCTTGATCTTC 20714
QY 556 TTTCTTCTGTGATGGTGTACTTCAGATGAATCCAGCCGTTATATATGGACCATGTA 615
Db 20713 TTTCTTCTGTGATGGTGTACTTCAGATGAATCCAGCCGTTATATATGGACCATGTA 20654
QY 616 GTAAACCTATCTGTGATCTGATAAGTGACATGGGTAACTTAGTTCATGCAACCAATAGTG 675
Db 20653 GTAAACCTATCTGTGATCTGATAAGTGACATGGGTAACTTAGTTCATGCAACCAATAGTG 20594
QY 676 AAGAAGATATTAATAAATTTATGCCAGATGATTTTAAGCTTAATGCGAGATTCCTTGATG 735
Db 20593 AAGAAGATATTAATAAATTTATGCCAGATGATTTTAAGCTTAATGCGAGATTCCTTGATG 20534
QY 736 GATTGATTATCTTCAGTGCAGATACTCCCTGTGTTTCTTCAACAGACCATGATAGTG 795
Db 20533 GATTGATTATCTTCAGTGCAGATACTCCCTGTGTTTCTTCAACAGACCATGATAGTG 20474
QY 796 ATACTGTCAAGAAACAACAGAAATGATATCAGTTCTGAAATACAAAATGAGAAATCGGAG 855
Db 20473 ATACTGTCAAGAAACAACAGAAATGATATCAGTTCTGAAATACAAAATGAGAAATCGGAG 20414
QY 856 GAATCAAGAAATTTGGGTATAAAGTAGATACAACTTTTCAGATTCCTATATTAATACAGTG 915
Db 20413 GAATCAAGAAATTTGGGTATAAAGTAGATACAACTTTTCAGATTCCTATATTAATACAGTG 20354
QY 916 GAACAGAAAATTTAAAAGATAAAAAGATCTTTAATCAGTTAGAAATCAATTTGTTATTTA 975
Db 20353 GAACAGAAAATTTAAAAGATAAAAAGATCTTTAATCAGTTAGAAATCAATTTGTTATTTA 20294
QY 976 ACATGTCTCTCTTTGACTCGACAAAGTTCCAAAATGTTTCATGCCAAAGCAAGCTAC 1035
Db 20293 ACATGTCTCTCTTTGACTCGACAAAGTTCCAAAATGTTTCATGCCAAAGCAAGCTAC 20234
QY 1036 AACACAAGAGCCAGCATGTGGATTACTAAAAGATGTTGGCTTAGTAAAGAGGAGTAG 1095
Db 20233 AACACAAGAGCCAGCATGTGGATTACTAAAAGATGTTGGCTTAGTAAAGAGGAGTAG 20174
QY 1096 ATGTGGCAGTCAATACTGCCGAGAAATGTTTAAAAGAGAGGCGCAAGCAAGTCTTTGA 1155
Db 20173 ATGTGGCAGTCAATACTGCCGAGAAATGTTTAAAAGAGAGGCGCAAGCAAGTCTTTGA 20114
QY 1156 CTGCGAGCCCTCCGAAAATGAAGATTTATGCTTTAAATGATTCAAAATCAAGAGATGAA 1215
Db 20113 CTGCGAGCCCTCCGAAAATGAAGATTTATGCTTTAAATGATTCAAAATCAAGAGATGAA 20054
QY 1216 ATTTCAAAATTAACCTGACTTTTCTCTTTCAGGAAGATAAGACTGTTTAAACCAATCTGCAC 1275
Db 20053 ATTTCAAAATTAACCTGACTTTTCTCTTTCAGGAAGATAAGACTGTTTAAACCAATCTGCAC 19994
```

QY 1276 AAGAGACTCAAAAGTTTACCTTAAGATATAGTAAATCCAAAGATTCCTCTTCAG 1335
 Db 1993 AAGAGACTCAAAAGTTTACCTTAAGATATAGTAAATCCAAAGATTCCTCTTCAG 1934
 QY 1336 CTTTACATGTTTCCAGTAAAGATGCGGCTCCTCATTTGCTGCTGCTTCCGCTGCGGT 1395
 Db 1993 CTTTACATGTTTCCAGTAAAGATGCGGCTCCTCATTTGCTGCTGCTTCCGCTGCGGT 19874
 QY 1396 CTATGCTGGATCATTAATTCAGAGTAAAGACAGGGGTGATTTTACCTCAGCATGAAC 1455
 Db 19873 CTATGCTGGATCATTAATTCAGAGTAAAGACAGGGGTGATTTTACCTCAGCATGAAC 19814
 QY 1456 ATAAAGATAATATACAGATGACGTGACTATACATGAAAGAAATACAGAAACAGTGTGTTTC 1515
 Db 19813 ATAAAGATAATATACAGATGACGTGACTATACATGAAAGAAATACAGAAACAGTGTGTTTC 19754
 QY 1516 TAGTGGGGAAACCATTCAGAGATGATCTTTTGAACACAGGAATAATGTAAGCATAC 1575
 Db 19753 TAGTGGGGAAACCATTCAGAGATGATCTTTTGAACACAGGAATAATGTAAGCATAC 19694
 QY 1576 TCCTTCAGTCAATTAATTCAGAGGATGGAAGACAGAAAGATAGATCTTCGACACAGCATAA 1635
 Db 19693 TCCTTCAGTCAATTAATTCAGAGGATGGAAGACAGAAAGATAGATCTTCGACACAGCATAA 19634
 QY 1636 TCAGAGTCAATTCAGAGGATGGAAGACAGAAAGATAGATCTTCGACACAGCATAA 1695
 Db 19633 TCAGAGTCAATTCAGAGGATGGAAGACAGAAAGATAGATCTTCGACACAGCATAA 19574
 QY 1696 GCGTTTCTGGCACTCATGTCGACAGTCTCTGATGTTGTAAGGTTTATTAAATCACT 1755
 Db 19573 GCGTTTCTGGCACTCATGTCGACAGTCTCTGATGTTGTAAGGTTTATTAAATCACT 19514
 QY 1756 TTTCAAGCAATCAATGATGGCAAGACTTAGATTAATCAATTAATGATCAAGGCGCAA 1815
 Db 19513 TTTCAAGCAATCAATGATGGCAAGACTTAGATTAATCAATTAATGATCAAGGCGCAA 19454
 QY 1816 AAGTGGCCCACTAATGATGATGCTGACAGTCTGATGCTTCTGACAGACATGATCTTC 1875
 Db 19453 AAGTGGCCCACTAATGATGATGCTGACAGTCTGATGCTTCTGACAGACATGATCTTC 19394
 QY 1876 AGACACTCAATCAATGATGCTTCTGACAGAAATGTAATGATCTCAAAATCGCAATCAATC 1935
 Db 19393 AGACACTCAATCAATGATGCTTCTGACAGAAATGTAATGATCTCAAAATCGCAATCAATC 19334
 QY 1936 AGATAGATATGAAGGCTTAGATGATGGAACATCAATTAATATATATTTCAATGCAGAG 1995
 Db 19333 AGATAGATATGAAGGCTTAGATGATGGAACATCAATTAATATATATTTCAATGCAGAG 19274
 QY 1996 CAGGAGCTATTTGGGAAAGTCACTGATTAATATAATTTGTAACACAGTTGATAACAAA 2055
 Db 19273 CAGGAGCTATTTGGGAAAGTCACTGATTAATATAATTTGTAACACAGTTGATAACAAA 19214
 QY 2056 ATACAATAAGAAATGGGCTTTCTTTAGAGAAAAAGACATATTCAGTTCAACAGGGT 2115
 Db 19213 ATACAATAAGAAATGGGCTTTCTTTAGAGAAAAAGACATATTCAGTTCAACAGGGT 19154
 QY 2116 TACCTACAGTAAAGTCTGAGATTAACAATCAATTAATCACTGCTGATTAATTAACAGTCAAT 2175
 Db 19153 TACCTACAGTAAAGTCTGAGATTAACAATCAATTAATCACTGCTGATTAATTAACAGTCAAT 19094
 QY 2176 CTGTTGAGGGCCAGACCTTAAGCAATTTAGCCCTTCCATCAAGAAACAGGAGTTCAA 2235
 Db 19093 CTGTTGAGGGCCAGACCTTAAGCAATTTAGCCCTTCCATCAAGAAACAGGAGTTCAA 19034
 QY 2236 AGGACCTGAATAGCCAGATGTTCCAGATACAAATGAAAGTGAACCCAGACAGCAGATA 2295
 Db 19033 AGGACCTGAATAGCCAGATGTTCCAGATACAAATGAAAGTGAACCCAGACAGCAGATA 18974
 QY 2296 CCGTTGTTCCCAATCACTTGTCTATAGATTTACAGCTGATCCACAGGTTAGCTTCAACT 2355
 Db 18973 CCGTTGTTCCCAATCACTTGTCTATAGATTTACAGCTGATCCACAGGTTAGCTTCAACT 18914
 QY 2356 CTAATTAATGATATAGAAAGTAATTTCTGAGGTTGATCTAGTTTCGTAACATGCAATG 2415

Db 18913 CTAATTAATGATATAGAAAGTAATTTCTGAAGGTGGATCTAGTTTCGTAACATGCAATG 18854
 QY 2416 AGATTCCTGCTACCTGAAACACTTCCAAAGAGGCTTGGTTTGGCCGCAAGACAGCCTA 2475
 Db 18853 AGATTCCTGCTACCTGAAACACTTCCAAAGAGGCTTGGTTTGGCCGCAAGACAGCCTA 18794
 QY 2476 CTTGGGTTCTCTGATTCAGAAAGCTCCAAACTGTATGAATGCAAGTCAAAATTTACTTTTA 2535
 Db 18793 CTTGGGTTCTCTGATTCAGAAAGCTCCAAACTGTATGAATGCAAGTCAAAATTTACTTTTA 18734
 QY 2536 CCAACCGGCGACACCATTCGCGAGCATGTGGAAAGTATTTTGT 2579
 Db 18733 CCAACCGGCGACACCATTCGCGAGCATGTGGAAAGTATTTTGT 18690
 RESULT 12
 AC008771 123169 bp DNA linear PRI 28-FEB-2001
 LOCUS Homo sapiens chromosome 5 clone CTD-2015H6, complete sequence.
 AC008771
 AC008771.4 GI:13162500
 VERSION HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 123169)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 123169)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 123169)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (28-FEB-2001) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 COMMENT On Feb 28, 2001 this sequence version replaced gi:7709299.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center
 www.shgc.stanford.edu
 Quality: Phrap Quality >=40 99.3% of Sequence;
 Estimated Total Number of Errors is 0.9.
 STS Content:
 SHGC-84992 GS3847
 WI-15051 G23648
 WI-17790 G24283.
 FEATURES
 Location/Qualifiers
 source
 1..123169
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CTD-2015H6"

ORIGIN

Query Match 34.0%; Score 2252.8; DB 9; Length 123169;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2257; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 316 CAGATGACACAGATTTATCTCCGAGATGTACAAAATGCATATGATTTCAACACTGCTCAG 375
 Db 43251 CAGATGACACAGATTTATCTCCAGATGTACAAAATGCATATGATTTCAACACTGCTCAG 43310
 QY 376 TTTCTTCAGAGTTGGCTTCTCCACAGCGAACTTCATGCTCCCAAGACCAAGAGTGGC 435
 Db 43311 TTTCTTCAGAGTTGGCTTCTCCACAGCGAACTTCATGCTCCCAAGACCAAGAGTGGC 43370

QY 436 TTAATAGTTGTGCTCATCAGAACCAAGCTATGGACCAATGAGATTCCTCCTGAATGAA 495
Db 43371 TTAATAGTTGTGCTCATCAGAACCAAGCTATGGACCAATGAGATTCCTCCTGAATGAA 43430
QY 496 AAACACCTCAAGGGACTTACTTCTATACAAATGAAAAAATGTAACAGGACTTGCCTTC 555
Db 43431 AAACACCTCAAGGGACTTACTTCTATACAAATGAAAAAATGTAACAGGACTTGCCTTC 43490
QY 556 TTTCTTCTGTGGATGGTGTGACTTACATGAAATCCAGCGTTATATATAGGGACGATGTA 615
Db 43491 TTTCTTCTGTGGATGGTGTGACTTACATGAAATCCAGCGTTATATATAGGGACGATGTA 43550
QY 616 GTAAACCTACTGTGATCTCATTAAGTCACATGGTAACTTAGTTTCATGCAACCAATAGTG 675
Db 43551 GTAAACCTACTGTGATCTCATTAAGTCACATGGTAACTTAGTTTCATGCAACCAATAGTG 43610
QY 676 AAGAAGATATTAATAAATATGCCAGATGATTTAAGTCTAATGAGATTCCTTGAATG 735
Db 43611 AAGAAGATATTAATAAATATGCCAGATGATTTAAGTCTAATGAGATTCCTTGAATG 43670
QY 736 GATTGGATTTATCTTCAGTGTGATGATCTCCCTGTGTTCTCAACAGACCATGATGATG 795
Db 43671 GATTGGATTTATCTTCAGTGTGATGATCTCCCTGTGTTCTCAACAGACCATGATGATG 43730
QY 796 ATACTGTGAGAGAACACAGATGATATCACTTCTGAAATACAAATAGAGAAATCGAG 855
Db 43731 ATACTGTGAGAGAACACAGATGATATCACTTCTGAAATACAAATAGAGAAATCGAG 43790
QY 856 GAATCAAGAAATGGGTATTAAGTAGATACACACTTTCAGATTCCTATATTAATGATG 915
Db 43791 GAATCAAGAAATGGGTATTAAGTAGATACACACTTTCAGATTCCTATATTAATGATG 43850
QY 916 GAACAGAAAAATTAAGAGATAAAGATCTTTAATCAGTTAGATCAATGTTGATTTTA 975
Db 43851 GAACAGAAAAATTAAGAGATAAAGATCTTTAATCAGTTAGATCAATGTTGATTTTA 43910
QY 976 ACATGTCATCTGCTTTCAGTCGACAAAGTTCCTCAAAATGTTTCATGCAAGCAAGCTAC 1035
Db 43911 ACATGTCATCTGCTTTCAGTCGACAAAGTTCCTCAAAATGTTTCATGCAAGCAAGCTAC 43970
QY 1036 AACACAGAGCCAGCCATGGAATTAATAAGATGTTGCTTAGTAAAGAGAGAGATG 1095
Db 43971 AACACAGAGCCAGCCATGGAATTAATAAGATGTTGCTTAGTAAAGAGAGAGATG 44030
QY 1096 ATGTGGCAGTCATAACTGCCGAGAAATGTTTAAAGAGAGGCGCAAGCAAGTCTTGA 1155
Db 44031 ATGTGGCAGTCATAACTGCCGAGAAATGTTTAAAGAGAGGCGCAAGCAAGTCTTGA 44090
QY 1156 CCTGCAGCTTCCGAAAAATGAGATTTATGCTTTAAATGATTCAAATTCAGAGATGAA 1215
Db 44091 CCTGCAGCTTCCGAAAAATGAGATTTATGCTTTAAATGATTCAAATTCAGAGATGAA 44150
QY 1216 ATTTCAAATACCTGACTTTTCTTTCAGGAAGATAAGACTGTTATATAAACAATCTGCAC 1275
Db 44151 ATTTCAAATACCTGACTTTTCTTTCAGGAAGATAAGACTGTTATATAAACAATCTGCAC 44210
QY 1276 AAGAAGACTCAAAAGTTTAGACCTTAAGGATATGATGTAATCCAGATTCCTCTTCAG 1335
Db 44211 AAGAAGACTCAAAAGTTTAGACCTTAAGGATATGATGTAATCCAGATTCCTCTTCAG 44270
QY 1336 CTTTACATGTTTCCAGTAAAGATGCGCTCCTCATTTGCTCTGCTTCTGCTGCTGGGT 1395
Db 44271 CTTTACATGTTTCCAGTAAAGATGCGCTCCTCATTTGCTCTGCTTCTGCTGCTGGGT 44330
QY 1396 CTATGTGGATCATTAATTAAGAAAGTAAAGCAGCGGTGATTTTTACCTCAGCATGAAC 1455
Db 44331 CTATGTGGATCATTAATTAAGAAAGTAAAGCAGCGGTGATTTTTACCTCAGCATGAAC 44390
QY 1456 ATAAAGATATATACAGATGCGATGATATACATGAAGAAATACAGACAGTGTGTTTC 1515
Db 44391 ATAAAGATATATACAGATGCGATGATATACATGAAGAAATACAGACAGTGTGTTTC 44450
QY 1516 TAGTGGGGGAACCATTCACAGAGAAATGATCTTTTGAACAGGAAAAATGTAAGAGCATAC 1575

Db 44451 TAGTGGGGGAACCATTCACAGAGAAATGATCTTTTGAACAGGAAAAATGTAAGAGCATAC 44510
QY 1576 TCCITTCAGTCATTAATTAAGAGGATGGAAGACAGAAAGATAGATCTTCACAGACAGTAA 1635
Db 44511 TCCITTCAGTCATTAATTAAGAGGATGGAAGACAGAAAGATAGATCTTCACAGACAGTAA 44570
QY 1636 TCAGAGCTGAGTCTTTGATGATGCTGACACAGTTCTACAGTTCTAGAAATCTCAAGAGG 1695
Db 44571 TCAGAGCTGAGTCTTTGATGATGCTGACACAGTTCTACAGTTCTAGAAATCTCAAGAGG 44630
QY 1696 GCGTTTCTGGCACTCATGTCCTCAGAGTCTTCTGATGTTGTGAAGGTTTATTAATACTT 1755
Db 44631 GCGTTTCTGGCACTCATGTCCTCAGAGTCTTCTGATGTTGTGAAGGTTTATTAATACTT 44690
QY 1756 TTTCAAGCAATGATATGATGCGCAAGACTTAGATTAATTAATGATGAAGCGCAA 1815
Db 44691 TTTCAAGCAATGATATGATGCGCAAGACTTAGATTAATTAATGATGAAGCGCAA 44750
QY 1816 AAAGTGGGCCACTAATTAATGATGCTGAACCTTGTAGTCCCTTTCTGACAGAACAGTATCTTC 1875
Db 44751 AAAGTGGGCCACTAATTAATGATGCTGAACCTTGTAGTCCCTTTCTGACAGAACAGTATCTTC 44810
QY 1876 AGACCACTAAACATAAAGTCTTTTGAAGAAATGTAATGACTCTTAATCGCAATGAATC 1935
Db 44811 AGACCACTAAACATAAAGTCTTTTGAAGAAATGTAATGACTCTTAATCGCAATGAATC 44870
QY 1936 AGATAGATATGAAGGCTTAGATGATGAAACATCAATAATATATATTTCAATGACAGAG 1995
Db 44871 AGATAGATATGAAGGCTTAGATGATGAAACATCAATAATATATATTTCAATGACAGAG 44930
QY 1996 CAGGAGCTATTGGGAAAGTCAATGATTAATTAATTTGTGAAACAGTTGATAACAAA 2055
Db 44931 CAGGAGCTATTGGGAAAGTCAATGATTAATTAATTTGTGAAATAGTTGATAACAAA 44990
QY 2056 ATACAATGAAATCGCCTTTCTTTAGGAGAAAAGCACTATTCCAGTTCAACAGGCT 2115
Db 44991 ATACAATGAAATCGCCTTTCTTTAGGAGAAAAGCACTATTCCAGTTCAACAGGCT 45050
QY 2116 TACCTACCAGTAAGTCTGAGATTAACAATCAATTAATGCTCTGATATTAACAGTCAAT 2175
Db 45051 TACCTACCAGTAAGTCTGAGATTAACAATCAATTAATGCTCTGATATTAACAGTCAAT 45110
QY 2176 CTGTTGGAGGGCCAGACCTTAAGCAATTTGTTAGCTTCCATCAAGAACAGAGTTCAA 2235
Db 45111 CTGTTGGAGGGCCAGACCTTAAGCAATTTGTTAGCTTCCATCAAGAACAGAGTTCAA 45170
QY 2236 AGGACTGTAATAAGCCAGATGTTCCAGATACAATAAGAAAGTGAACCCAGCACAGAGATA 2295
Db 45171 AGGACTGTAATAAGCCAGATGTTCCAGATACAATAAGAAAGTGAACCCAGCACAGAGATA 45230
QY 2296 CCGTTGTTCCAAATCACTTGTGCTATAGATTCACAGCTGATCCACAGGTTAGCTTCAACT 2355
Db 45231 CCGTTGTTCCAAATCACTTGTGCTATAGATTCACAGCTGATCCACAGGTTAGCTTCAACT 45290
QY 2356 CTAAATTCATTCATATAGAAAGTAAATCTGAAAGTGGATCTAGTTTCGTAACCTGCAATG 2415
Db 45291 CTAAATTCATTCATATAGAAAGTAAATCTGAAAGTGGATCTAGTTTCGTAACCTGCAATG 45350
QY 2416 AAGATTCTGTACTGAAAAACACTTGCACAAAGAGGCTTGTTTGGGCCAGAAACAGCCTA 2475
Db 45351 AAGATTCTGTACTGAAAAACACTTGCACAAAGAGGCTTGTTTGGGCCAGAAACAGCCTA 45410
QY 2476 CTGCGTTCTGATTCAGAGCTCCAAAGCTGATGAACTGCAAGTCAAAATTAATTTA 2535
Db 45411 CTGCGTTCTGATTCAGAGCTCCAAAGCTGATGAACTGCAAGTCAAAATTAATTTA 45470
QY 2536 CCAAAAGCGGACACCAATGCGGAGCATGTGGGAAAAGTATTTTGT 2579
Db 45471 CCAAAAGCGGACACCAATGCGGAGCATGTGGGAAAAGTATTTA 45514

RESULT 13

AC018764/c
 LOCUS AC018764 126052 bp DNA linear PRI 13-MAR-2003
 DEFINITION Homo sapiens chromosome 5 clone CTD-2327L5, complete sequence.
 ACCESSION AC018764
 VERSION AC018764.8 GI:28933541
 HTG.
 KEYWORDS Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 126052)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 Direct Submission
 Unpublished
 2 (bases 1 to 126052)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (19-DEC-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 3 (bases 1 to 126052)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 Direct Submission
 Submitted (01-JUL-2000) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 4 (bases 1 to 126052)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 Direct Submission
 Submitted (09-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 5 (bases 1 to 126052)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 Direct Submission
 Submitted (30-JAN-2003) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 6 (bases 1 to 126052)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 Direct Submission
 Submitted (13-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 On Mar 13, 2003 this sequence version replaced gi:28144392.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center
 www.sbgc.stanford.edu
 Quality: Phrap Quality >=40 99.6% of Sequence;
 Estimated Total Number of Errors is 0.6.
 NOTE: Bacterial transposon excised at 89327.
 Location/Qualifiers
 1..126052
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CTD-2327L5"

FEATURES
 source

ORIGIN
 Query Match 28.8%; Score 1910.4; DB 9; Length 126052;
 Best Local Similarity 99.9%; Pred. No. 2e-298;
 Matches 1922; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 4709 GGTGAAATTCAGGAGATCTGAAGCCAACTTCTGCCTCAGCAATATCTAAATGATCT 4768
 DB 97844 GGTGAAATTCAGGAGATCTGAAGCCAACTTCTGCCTCAGCAATATCTAAATGATCT 97785
 QY 4769 TGATAGTCTCTGATACCTGATCCATGGTGGGACCTCCAACTCTAGTTTACCATTAGA 4828
 DB 97784 TGATAGTCTCTGATACCTGATCCATGGTGGGACCTCCAACTCTAGTTTACCATTAGA 97725
 QY 4829 AATAGAATTTAGTGTCTTTTCAATATAGAACATCTTTTGTAGTGAAGAAATCTGCCATATTA 4888
 DB 97724 AATAGAATTTAGTGTCTTTTCAATATAGAACATCTTTTGTAGTGAAGAAATCTGCCATATTA 97665
 QY 4889 CATATTGCAACTAATTTGTTTAAACTAACTCCAGCACTAAAGCTGAAGTGCACCAACA 4948

DB 97664 CATATTGCAACTAATTTGTTTAAACTAACTCCAGCACTAAAGCTGAAGTGCACCAACA 97605
 QY 4949 CTAAAGATATAATATGCTGATTTTGAACACACATAAGCTTTGCTCTTTTAGGAGGAAT 5008
 DB 97604 CTAAAGATATAATATGCTGATTTTGAACACACATAAGCTTTGCTCTTTTAGGAGGAAT 97545
 QY 5009 GATCTTTTCAATCAATTAGCAATATTTAAATATCTAAAAATTTAAGAGATCCATCTT 5068
 DB 97544 GATCTTTTCAATCAATTAGCAATATTTAAATATCTAAAAATTTAAGAGATCCATCTT 97485
 QY 5069 TCTGTAGCTTTACAAATTAATTTAAGTACTAAAGACAGGATTTCTTTTAAGAAATTTA 5128
 DB 97484 TCTGTAGCTTTACAAATTAATTTAAGTACTAAAGACAGGATTTCTTTTAAGAAATTTA 97425
 QY 5129 TAGCAATTTACTGTGTATTTAAATCTAAGCCAAAGTATCTGCCTTAGGTATACCTCTT 5188
 DB 97424 TAGCAATTTACTGTGTATTTAAATCTAAGCCAAAGTATCTGCCTTAGGTATACCTCTT 97365
 QY 5189 TAGCCAAATATGATTTAATGAAGGCTCTTTTCAAGATGAACCTTATGAAGGAAATATC 5248
 DB 97364 TAGCCAAATATGATTTAATGAAGGCTCTTTTCAAGATGAACCTTATGAAGGAAATATC 97305
 QY 5249 TGCCTTGTGTATATGCCAGTTAGATCTGGTTTCTAAAGTCTGTCAAAATTTGTAATTCAG 5308
 DB 97304 TGCCTTGTGTATATGCCAGTTAGATCTGGTTTCTAAAGTCTGTCAAAATTTGTAATTCAG 97245
 QY 5309 TGGCAAAAAACAGTTTGGAGTCTTAGACTTATAATCTTTGAATAAACTGATAACT 5368
 DB 97244 TGGCAAAAAACAGTTTGGAGTCTTAGACTTATAATCTTTGAATAAACTGATAACT 97185
 QY 5369 TATTTGTATAATTCAGATGAAGGACCTCTCCATAATTTAGATAAACTCTTTTGGATTAT 5428
 DB 97184 TATTTGTATAATTCAGATGAAGGACCTCTCCATAATTTAGATAAACTCTTTTGGATTAT 97125
 QY 5429 AATCAGAATTTTGCCTTTTCTTCTCAAAATTTATACATATGTATGTATATATATCCAC 5488
 DB 97124 AATCAGAATTTTGCCTTTTCTTCTCAAAATTTATACATATGTATGTATATATATCCAC 97065
 QY 5489 ATATATAGTTTCCCTGATTAATGAATTAATAATTTTCCGGTCTTCAGGACCTTTT 5548
 DB 97064 ATATATAGTTTCCCTGATTAATGAATTAATAATTTTCCGGTCTTCAGGACCTTTT 97005
 QY 5549 TGCCTTCTATATTTAAGTATATTTTATAGCAAGAACATATTTCTGAATGTTTATAAA 5608
 DB 97004 TGCCTTCTATATTTAAGTATATTTTATAGCAAGAACATATTTCTGAATGTTTATAAA 96945
 QY 5609 TCTTTAATAATTTATATGATAGTAATATTTTGTATCAAAATGCAATGCAATATTTTCTC 5668
 DB 96944 TCTTTAATAATTTATATGATAGTAATATTTTGTATCAAAATGCAATGCAATATTTTCTC 96886
 QY 5669 CTTTCTCTCCAACTATACCTGATTTTACACTCTTAAGAGTACCTGACGACGGGCCA 5728
 DB 96885 CTTTCTCTCCAACTATACCTGATTTTACACTCTTAAGAGTACCTGACGACGGGCCA 96826
 QY 5729 GATGACCTTTGAAGTATGATTTAGCAATAAATGAAGCCTGAAACAGGTTTCTTTTAC 5788
 DB 96825 GATGACCTTTGAAGTATGATTTAGCAATAAATGAAGCCTGAAACAGGTTTCTTTTAC 96766
 QY 5789 TTCCACTTTATCCCTTAGAAATTTCTTGGCAACTTCGCATATTTTTCATGACACCTGGTGT 5848
 DB 96765 TTCCACTTTATCCCTTAGAAATTTCTTGGCAACTTCGCATATTTTTCATGACACCTGGTGT 96706
 QY 5849 ATAAGTATAAATTTAAATGAACATAATTTTGTGATATTTTAAATCTTTTATATGATG 5908
 DB 96705 ATAAGTATAAATTTAAATGAACATAATTTTGTGATATTTTAAATCTTTTATATGATG 96646
 QY 5909 TTATTTTATACAGGATATTAACATAGTAAATCTTATGATTTTCAAAATTTGTTACAG 5968
 DB 96645 TTATTTTATACAGGATATTAACATAGTAAATCTTATGATTTTCAAAATTTGTTACAG 96586
 QY 5969 AGCTTTCTCTTTTACTTCAAAACAGCAAAAAAGTGGGGGCATATTTGTAGTCTCTGCTATT 6028

Best Local Similarity 98.4%; Pred. No. 3.5e-286;
Matches 1896; Conservative 0; Mismatches 20; Indels 10; Gaps 4;

QY	4709	GGTTCAATTTTCAGGAGGATCTGAAGGCAACTTCTGCTCAGCATATCTAAATGATCT	4768
DB	626	GGTGAATTTTCAGGAGGATCTGAAGGCAACTTCTGCTCAGCATATCTAAATGATCT	685
QY	4769	TGATAGTCTCTGATACCTGATGATCCATGCTGGGACCTCCAACTCTAGTTTACCATTAGA	4828
DB	686	TCATAGTCTCTGATACCTGATGATCCATGCTGGGACCTCCAACTCTAGTTTACCATTAGA	745
QY	4829	AATGAATTTAGTGTCTTTCATTAATAGACATCTTTTATAGGAAGAAATGCGCCATATTA	4888
DB	746	AATGAATTTAGTGTCTTTCATTAATAGACATCTTTTATAGGAAGAAATGCGCCATATTA	805
QY	4889	CATATTGCAACCTAATTTGTTAAACCTAACTCCAGCCTAAAGCTGAATGCCCAAAACA	4948
DB	806	CATATTGCAACCTAATTTGTTAAACCTAACTCCAGCCTAAAGCTGAATGCCCAAAACA	865
QY	4949	CTAAAGTATTAATATGCTGATTTTGAACACATAAGCTTTGCTCTTTAGGAGGAAT	5008
DB	866	CTAAAGTATTAATATGCTGATTTTGAACACATAAGCTTTGCTCTTTAGGAGGAAT	923
QY	5009	GATCTTTTCAATCATTAAGCAATATTTAAATATCTAAATTTTAAAGAGATCCATACCT	5068
DB	924	GATCTTTTCAATCATTAAGCAATATTTAAATATCTAAATTTTAAAGAGATCCATACCT	983
QY	5069	TCTGAGCTTTTCAATTAATTAAGTACTAAAGACAGAGATTTCTTTTAAAGAAATTA	5128
DB	984	TCTGAGCTTTTCAATTAATTAAGTACTAAAGACAGAGATTTCTTTTAAAGAAATTA	1043
QY	5129	TAGCATTTTACTGTGTTATTTAAATGCTAAAGCAAGATCTGCACTTTAGGTATACCTCT	5188
DB	1044	TAGCATTTTACTGTGTTATTTAAATGCTAAAGCAAGATCTGCACTTTAGGTATACCTCT	1103
QY	5189	TATGCCAATAATGATTTTAAATGAAGGCTCTTTTCAAGATGTAACCTTATGAAGGAATATC	5248
DB	1104	TATGCCAATAATGATTTTAAATGAAGGCTCTTTTCAAGATGTAACCTTATGAAGGAATATC	1163
QY	5249	TGCTTTTGATATATGCCAGTTAGATATCTGTTTCTAAAGTCTGCAATTTGATTTTCAG	5308
DB	1164	TGCTTTTGATATATGCCAGTTAGATATCTGTTTCTAAAGTCTGCAATTTGATTTTCAG	1223
QY	5309	TGGCACAATAATGATTTTGAAGGCTCTTGAAGTCTTAAATTTCTTGAATAAACTGATAACT	5368
DB	1224	TGGCACAATAATGATTTTGAAGGCTCTTGAAGTCTTAAATTTCTTGAATAAACTGATAACT	1283
QY	5369	TATTTGATATATGAGTGGAGACTTACCTCCATAATAGATAAATCTTTTGGATTAT	5428
DB	1284	TATTTGATATATGAGTGGAGACTTACCTCCATAATAGATAAATCTTTTGGATTAT	1343
QY	5429	AATCAGAAATTTGCTTTTCTTCTCAAAATTTATACATATGATGATTTATATATCCAC	5488
DB	1344	AATCAGAAATTTGCTTTTCTTCTCAAAATTTATACATATGATGATTTATATATCCAC	1403
QY	5489	ATATATAGTTTCCCTGATTAATTTGATATTAATTAATTTGCGGCTCTTCAAGACTTTT	5548
DB	1404	ATATATAGTTTCCCTGATTAATTTGATATTAATTAATTTGCGGCTCTTCAAGACTTTT	1463
QY	5549	TGCTTCTATATTTAAGTATATTTTATAGCAAGAACATATTTCTGAA--TGTTTTATA	5606
DB	1464	TGCTTCTATATTTAAGTATATTTTATAGCAAGAACATATTTCTGAA--TGTTTTATA	1523
QY	5607	AATCTTTAATTAATTTATGATGATTAATTTTGTATCAATATGCAATTTTCTTCC	5666
DB	1524	AATCTTTAATTAATTTATGATGATTAATTTTGTATCAATATGCAATTTTCTTCC	1582
QY	5667	TCCTTCTTCCCAACTATACCTGATTTTACCACTTCTTAAGAGTACTGACGAGGCG	5726
DB	1583	TCCTTCTTCCCAACTATACCTGATTTTACCACTTCTTAAGAGTACTGACGAGGCG	1642
QY	5727	CAGATGACCTTTGAAGTAGTCAATTTATGATAGCAATAATTAAGCTGCAAAACAGGTTTTT	5786

DB	1643	CAGATGACCTTTGAAGTAGTCAATTTATGATAGCAATAATTAAGCTGAAACAGGTTTTTT	1702
QY	5787	ACTTCCACATTTAATCCTTAGAAATTTCTTGGCAACTTTCGCATATTTTCATTGACACTGGT	5846
DB	1703	ACTTCCACATTTAATCCTTAGAAATTTCTTGGCAACTTTCGCATATTTTCATTGACACTGGT	1762
QY	5847	GTATAAGTATAATTTTAAATGAACATAATTAATCTTTTGGCATATTTTAAATCTTTATATGGT	5906
DB	1763	GTATAAGTATAATTTTAAATGAACATAATTAATCTTTTGGCATATTTTAAATCTTTATATGGT	1822
QY	5907	AGTTATTTTAAACAGGATATTAACATAAGTTAAATCTTATGATTTTAAATTTGTTTAC	5966
DB	1823	AGTTATTTTAAACAGGATATTAACATAAGTTAAATCTTATGATTTTAAATTTGTTTAC	1882
QY	5967	AGAGCTTCTCTTTTCAACAGCAAGAAAAGTGGGGGGCATATTTAGTCTGTCAT	6026
DB	1883	AGAGCTTCTCTTTTCAACAGCAAGAAAAGTGGGGGGCATATTTAGTCTGTCAT	1942
QY	6027	TTAAGTTATGTAATAAATTTTAAATCAATTAATTTTAAATCTTATGATTTTAAATTTGTTTAC	6086
DB	1943	TTAAGTTATGTAATAAATTTTAAATCAATTAATTTTAAATCTTATGATTTTAAATTTGTTTAC	2002
QY	6087	TATGTTTTTGTATCAAAAACACTCATATATTTTCAAGAAAAAGAAATTTATGTTAAATAGCC	6146
DB	2003	TATGTTTTTGTATCAAAAACACTCATATATTTTCAAGAAAAAGAAATTTATGTTAAATAGCC	2062
QY	6147	CTGTTTTAAGAAAAATTTTATGAAGCATCTCAAGCTTGAAGTCAAGTCAAGTATTAAC	6206
DB	2063	CTGTTTTAAGAAAAATTTTATGAAGCATCTCAAGCTTGAAGTCAAGTCAAGTATTAAC	2122
QY	6207	TCAGACTGAGGCTCTCAAGCTTAGGAGAGACTGAGAAATTTTAAATCAGTTTGGGCATATAG	6266
DB	2123	TCAGACTGAGGCTCTCAAGCTTAGGAGAGACTGAGAAATTTTAAATCAGTTTGGGCATATAG	2182
QY	6267	TTTGAGCTGAATCAGTCTGTAGTACTTAGCCAAAGACAAATTTGGAGAGAAATATCAGCC	6326
DB	2183	TTTGAGCTGAATCAGTCTGTAGTACTTAGCCAAAGACAAATTTGGAGAGAAATATCAGCC	2242
QY	6327	TTCTGGAAGTACTCTTCTGAAACAATGTAAGTGTGCGAGATATCAATAAATAGGCA	6386
DB	2243	TTCTGGAAGTACTCTTCTGAAACAATGTAAGTGTGCGAGATATCAATAAATAGGCA	2297
QY	6387	ACCTGTTATAATTTGTGAAATTTTAAATGTTGTAAGATGAAACAAATTTGCATATCAA	6446
DB	2298	ACCTGTTATAATTTGTGAAATTTTAAATGTTGTAAGATGAAACAAATTTGCATATCAA	2357
QY	6447	ACCCAATTTATGTTTTCTAAATATATGATGATTTCTGCCATGTAAGTATTTGAACAGT	6506
DB	2358	ACCCAATTTATGTTTTCTAAATATATGATGATTTCTGCCATGTAAGTATTTGAACAGT	2417
QY	6507	CTTAAATAACCAATGGTAGAGGCTCTTCCATGATGGGACGCTTTTGGATTTTGTTC	6566
DB	2418	CTTAAATAACCAATGGTAGAGGCTCTTCCATGATGGGACGCTTTTGGATTTTGTTC	2477
QY	6567	ATAAATCTCTACATCTCAATTAATTTGGAATTTATGTCCTGAACTTTGGGAGCACAAT	6626
DB	2478	ATAAATCTCTACATCTCAATTAATTTGGAATTTATGTCCTGAACTTTGGGAGCACAAT	2537
QY	6627	TGAACT 6632	
DB	2538	TGAAT 2543	

RESULT 15
AK090896

LOCUS Homo sapiens cDNA FLJ33577 fis, clone BRAY2011480.
DEFINITION AK090896
ACCESSION AK090896
VERSION AK090896.1 GI:21749142
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1

REFERENCE
AUTHORS
Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,
Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Irie,R.,
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamanoto,J., Isono,Y.,
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
Nagahara,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2622)
Isogai,T. and Yamamoto,J.
Direct Submission
TITLE
JOURNAL
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatori, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
Location/Qualifiers
1..2622
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="BRAM2011480"
/tissue_type="amygdala"
/clone_lib="BRAM2"
/note="Cloning vector: pME18SFL3"

ORIGIN

Query Match 25.2%; Score 1668.8; DB 9; Length 2622;
Best Local Similarity 99.8%; Pred. No. 4.8e-259;
Matches 1681; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 4709 GGTGTAATTCAGCAGGATCTGAAGGCCAACTTCGCTCAGCATTTCTAAATGATCT 4768
Db 940 GGTGTAATTCAGCAGGATCTGAAGGCCAACTTCGCTCAGCATTTCTAAATGATCT 999

Qy 4769 TGATAGTGCTCTGATACCTGTGATCCATGGTGGACCTCCAACTCTAGTTTACCATTAGA 4828
Db 1000 TGATAGTGCTCTGATACCTGTGATCCATGGTGGACCTCCAACTCTAGTTTACCATTAGA 1059

Qy 4829 AATAGAATTAGTGTCTTTTCATTATAGAACATCTTTTTTAGTGAAGAATGTGCCATTATA 4888
Db 1060 AATAGAATTAGTGTCTTTTCATTATAGAACATCTTTTTTAGTGAAGAATGTGCCATTATA 1119

Qy 4889 CATATTGCAACTAATTTGTTTAAACTTAACCTCCAGCATTAAGCTGAAATGCCCAACA 4948
Db 1120 CATATTGCAACTAATTTGTTTAAACTTAACCTCCAGCATTAAGCTGAAATGCCCAACA 1179

Qy 4949 CTAAGAAGTATAAATGCTGATTTTGAACACATAGCTTTTCTCTTTAGGCAGGAAT 5088
Db 1180 CTAAGAAGTATAAATGCTGATTTTGAACACATAGCTTTTCTCTTTAGGCAGGAAT 1239

Qy 5009 GATCTTTCAAATCATTAGCACAATATTAAATATCTAAAATTTAGAGATCCATCTT 5068
Db 1240 GATCTTTCAAATCATTAGCACAATATTAAATATCTAAAATTTAGAGATCCATCTT 1299

Qy 5069 TCTGTAGCTTTTACAAATTAATTTAGTACTAAAAAGACAAGGATTTCTTTAAGAAATTTA 5128
Db 1300 TCTGTAGCTTTTACAAATTAATTTAGTACTAAAAGACAAGGATTTCTTTAAGAAATTTA 1359

Qy 5129 TAGCATTTACTGTGTTATTTAAATGCTAAAGCAAAGTATCTGCATTTAGGTATATCTCTT 5188
Db 1360 TAGCATTTACTGTGTTATTTAAATGCTAAAGCAAAGTATCTGCATTTAGGTATATCTCTT 1419

[illegible]

Search completed: May 18, 2004, 02:10:37
Job time : 16485.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 09:46:50 ; Search time 1526.9 Seconds
(without alignments)

18451.753 Million cell updates/sec

Title: US-09-744-167-3

Perfect score: 6632

Sequence: 1 actccggcgccgggtagctc.....ttggaggccattttgaagt 6632

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 1 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq 29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1980s:*
3: Geneseqn2000s:*
4: Geneseqn2000s:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6632	100.0	6632	3 AAZ50069	Aaz50069 Human Sma
2	5290	79.8	5402	6 ABK11709	Abk11709 DNA encod
3	1514	22.8	1572	3 AAC98041	Aac98041 Human col
4	850	12.8	969	6 ABL89832	AbL89832 Human pol
5	665.4	10.0	693	3 AAA44354	Aaa44354 Human sec
6	580.2	8.7	701	6 ABQ57538	Abq57538 Human col
7	560.6	8.5	4839	3 AAZ50068	Aaz50068 Human Sma
8	556.2	8.4	1977	9 ADC64443	Adc64443 Human SAR
9	500	7.5	4823	3 AAZ50070	Aaz50070 Xenopus S
10	472.4	7.1	475	8 ACH16512	Ach16512 Human adu
11	444	6.7	455	5 AAH81649	Aah81649 Human dif
12	389.4	5.9	391	8 ACH20104	Ach20104 Human adu
13	286	4.3	317	4 AAH35041	Aah35041 Human col
14	199.6	3.0	393	7 ABX48202	Abx48202 Bovine ES
15	191.8	2.9	4068	4 ABL28217	AbL28217 Drosophil
16	187.4	2.8	277	2 AAT23367	Aat23367 Human gen
17	181.6	2.7	289	2 AAV88493	Aav88493 EST clone
18	180.8	2.7	6784	4 ABL28216	AbL28216 Drosophil
19	172	2.6	556	4 AAK11740	Aak11740 Human bra
20	168	2.5	168	4 AAK24333	Aak24333 Human bra
21	124	1.9	387	6 ABN93507	Abn93507 Gene #5 u
22	116.8	1.8	576	5 ABV54556	Abv54556 Human pro
23	105.2	1.6	8056	7 ABZ10246	Abz10246 Haematopo

24	101.4	1.5	8056	7	ABZ10246	Abz10246	Haematopo
c	25	98.8	1.5	8056	7	ABZ10100	Haematopo
	26	95	1.4	888	7	ACA56987	Human adi
	27	93.6	1.4	255	5	AAH82068	Rat diffe
c	28	86.6	1.3	8056	7	ABZ10100	Haematopo
	29	85.4	1.3	254	3	AAC27984	Human sec
	30	82	1.2	11805	6	ABL33748	Human imm
c	31	76	1.1	251	3	AAC07640	Human sec
	32	75.6	1.1	7597	6	ABL33013	Human imm
	33	75.2	1.1	7657	4	AAS45477	Chemical
c	34	75.2	1.1	7657	6	ABL34022	Human imm
	35	75.2	1.1	115218	7	ACA64845	Human HNR
	36	75	1.1	11812	4	AAS45502	Chemical
c	37	75	1.1	11812	4	AAS46742	Tumour su
	38	75	1.1	11812	6	ABL34119	Human imm
	39	75	1.1	11812	6	ABK28432	DNA trans
c	40	74	1.1	61020	4	AAS46788	Tumour su
	41	73.4	1.1	115218	7	ACA64845	Human HNR
	42	73.2	1.1	5689	4	AAS45384	Chemical
c	43	73.2	1.1	5689	4	AAS46426	Tumour su
	44	73.2	1.1	5689	6	ABK28426	DNA trans
	45	73	1.1	5930	6	ABL32517	Human imm

ALIGNMENTS

RESULT 1

AAZ50069
ID AAZ50069 standard; cDNA; 6632 BP.

XX AAZ50069;

DT 04-MAY-2000 (first entry)

DE Human Smad Anchor for Receptor Activation protein-2 encoding cDNA.

XX Smad Anchor for Receptor Activation protein; hSARA2; human;

KW transforming growth factor-beta; TGF-beta; bone morphogenetic protein;

KW BMP; activin; anti-inflammatory; cytostatic; antiarthritic; vulnery;

KW TGF-beta modulator; wound healing; scarring; arthritis; immune response;

KW inflammatory response; tumour progression; cell proliferation; fibrosis;

KW fibrogenesis; tissue morphogenesis; tissue damage; ss.

XX Homo sapiens.

XX Location/Qualifiers

PH Key

FT CDS

FT /tag= a

FT /product= "Human SARA-2 protein"

FT /note= "Binds to receptor regulated Smad proteins"

FT /transl_except= (pos:336..338, aa:Gln)

FT /transl_except= (pos:822..824, aa:Thr)

XX WO200005360-A1.

XX C3-FEB-2000.

XX 20-JUL-1999; 99WO-CA000656.

XX 20-JUL-1998; 98CA-02237701.

XX 10-DEC-1998; 98CA-02253647.

XX (HSCR-) HSC RES & DEV LP.

XX Wraha JL;

XX WPI; 2000-182691/16.

XX P-PSDB; AAY44750.

XX New Smad Anchor for Receptor Activation (SARA) polypeptides, useful for

XX developing agents for treating e.g. wounds, arthritis, fibrosis,

XX inflammation, tumors, fibrogenesis or tissue damage.

XX Claim 8; Page 53-55; 93pp; English.

XX The present sequence is the cDNA encoding the human Smad Anchor for

CC Receptor Activation protein-2 (hsARA2), a TGF-beta modulator. This has

CC been identified in the EST clone, KIA0305 and is related to hSARA1. The

CC hSARA proteins bind to receptor-regulated Smad proteins and ensures its

CC appropriate localisation for activation by a type I receptor of a TGF-

CC beta, activin or bone morphogenetic protein (BMP) signalling pathway.

CC SARA proteins has anti-inflammatory, cytoskeletal, anti-arthritis and

CC vulnery activity. The SARA protein is useful for developing agents for

CC prevention or treatment of disorders involving TGF-beta superfamily

CC member signalling pathway. Such agents can be used for wound healing,

CC scarring, arthritis and fibrosis (e.g., liver and kidney), in modulating

CC inflammatory and immune responses, tumour progression, cell proliferation

CC or fibrogenesis and in tissue morphogenesis. It is also used for

CC protecting, restoring and regenerating tissues after tissue damage

XX Sequence 6632 BP; 2197 A; 1123 C; 1253 G; 2059 T; 0 U; 0 Other;

QY Query Match 100.0%; Score 6632; DB 3; Length 6632;

PS Best Local Similarity 100.0%; Pred. No. 0;

XX Matches 6632; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACTCCCGCGGGGTAGTCTTCACTCTCAGCGGAGCTGCTGTCGAGTTCCTCCAAAAG 60

DB 1 ACTCCCGCGGGGTAGTCTTCTCTCTCTCAGCGGAGCTGCTGTCGAGTTCCTCCAAAAG 60

QY 61 CTCGCGAGGGGTAGGAGGTGATCTCATCCATTAAACAGCTGTGTGTCAGTTCCTCC 120

DB 61 CTCGCGAGGGGTAGGAGGTGATCTCATCCATTAAACAGCTGTGTGTCAGTTCCTCC 120

QY 121 AAATCTTATCTATCTCAGACTTCTCTCTCGATTCAGATTCCTATATTCAGCTGCCTT 180

DB 121 AAATCTTATCTATCTCAGACTTCTCTCTCGATTCAGATTCCTATATTCAGCTGCCTT 180

QY 181 TTGGATATCTCTCCAGGATGTTCTCAAGGCATCAAGAAATTAATCTGAATAGTCTG 240

DB 181 TTGGATATCTCTCCAGGATGTTCTCAAGGCATCAAGAAATTAATCTGAATAGTCTG 240

QY 241 CAGGTAGATGACAGTATTTTAAAGCAGCTGTCTGAGTTCGACAAAGTCTCTGATG 300

DB 241 CAGGTAGATGACAGTATTTTAAAGCAGCTGTCTGAGTTCGACAAAGTCTCTGATG 300

QY 301 ATTTTGAACAGAACCCAGATGAACAGATTTATCTCGCAGATGTAACAAATGCATATG 360

DB 301 ATTTTGAACAGAACCCAGATGAACAGATTTATCTCGCAGATGTAACAAATGCATATG 360

QY 361 CTAAACATGTCTCAGTTCCTCAGATGGTCTCTCTCAGCGAACTTCATGTCTCCAA 420

DB 361 CTAAACATGTCTCAGTTCCTCAGATGGTCTCTCTCAGCGAACTTCATGTCTCCAA 420

QY 421 AAGACCAAGAGTGGTAAATAGTGTGCTCATCAGAAACAGCTATGGAACAAATGAGA 480

DB 421 AAGACCAAGAGTGGTAAATAGTGTGCTCATCAGAAACAGCTATGGAACAAATGAGA 480

QY 481 GTTCCCTGAATGAAAAACACTCAAGGACTTACTTCTATACAAATGAAAAAATGTA 540

DB 481 GTTCCCTGAATGAAAAACACTCAAGGACTTACTTCTATACAAATGAAAAAATGTA 540

QY 541 CAGGACTGATCTCTCTCTCTGATGCTGATCTTCTCAGATGAATCCAGCGGTAT 600

DB 541 CAGGACTGATCTCTCTCTCTGATGCTGATCTTCTCAGATGAATCCAGCGGTAT 600

QY 601 ATATGGGACGATGTAGTAAACCTATCTGTGATCTGATAAGTGACATGGTAACTTAGTTC 660

DB 601 ATATGGGACGATGTAGTAAACCTATCTGTGATCTGATAAGTGACATGGTAACTTAGTTC 660

QY 661 ATGCACCAATAGTGAAGAGATATTAATAATTTATGCCAGATGATTTAAGTCTAATG 720

DB 661 ATGCACCAATAGTGAAGAGATATTAATAATTTATGCCAGATGATTTAAGTCTAATG 720

QY 721 CAGATTCTTGATTGGATTGGATTATCTTCAAGTGTGAGATCTCTGATCTCTCTCA 780

DB 721 CAGATTCTTGATTGGATTGGATTATCTTCAAGTGTGAGATCTCTCTGTTCTTCAA 780

QY 781 CAGACCATGATAGTGTACTGTGAGAAACAACAGAAATGATATCAGTTCTGAAATACAAA 840

DB 781 CAGACCATGATAGTGTACTGTGAGAAACAACAGAAATGATATCAGTTCTGAAATACAAA 840

QY 841 ATAGAGAAATCGAGAGATCAAGAAATGCGGTATTAAGTAGATACAACTTTCAGATT 900

DB 841 ATAGAGAAATCGAGAGATCAAGAAATGCGGTATTAAGTAGATACAACTTTCAGATT 900

QY 901 CCTATATTTACAGTGTGAACAGAAAATTTAAAGATATAAAGATCTTAAATCAGTTAGAT 960

DB 901 CCTATATTTACAGTGTGAACAGAAAATTTAAAGATATAAAGATCTTAAATCAGTTAGAT 960

QY 961 CAATTCTGATTTTAAATGATCATCTGTCTGACTCGACAAAGTTCCTCAAAATGTTTCATG 1020

DB 961 CAATTCTGATTTTAAATGATCATCTGTCTGACTCGACAAAGTTCCTCAAAATGTTTCATG 1020

QY 1021 CCAGACAGAGCTTACAAACAGAGCCAGCCATGTGGATTACTTAAAGATGTGCTTAG 1080

DB 1021 CCAGACAGAGCTTACAAACAGAGCCAGCCATGTGGATTACTTAAAGATGTGCTTAG 1080

QY 1081 TAAAGAGGAGATAGATGTGGCAGTCATAACTGCCGCGAGATGTTTAAAGAGAGAGGCA 1140

DB 1081 TAAAGAGGAGATAGATGTGGCAGTCATAACTGCCGCGAGATGTTTAAAGAGAGAGGCA 1140

QY 1141 AGCAAGTCTTGTACCTGCAGCCTTCGAAAATGAAGATTTATGCTTAATGATTTCAA 1200

DB 1141 AGCAAGTCTTGTACCTGCAGCCTTCGAAAATGAAGATTTATGCTTAATGATTTCAA 1200

QY 1201 ATTCAAGAGATCAAAATTTCAATTAACCTGACTTTTCTTTCAGGAAGATGAAGTCTGTA 1260

DB 1201 ATTCAAGAGATCAAAATTTCAATTAACCTGACTTTTCTTTCAGGAAGATGAAGTCTGTA 1260

QY 1261 TAAACAAATCTGCAAGAGAGCTCAAAAGTTTAGACTTTAGACTTTAGGATTAATGTAATCC 1320

DB 1261 TAAACAAATCTGCAAGAGAGCTCAAAAGTTTAGACTTTAGACTTTAGGATTAATGTAATCC 1320

QY 1321 AAGATTCCTCTTTCAGCTTTTACATGTTTCCAGTAAAGATGTGCGCTCTCATGCTGTC 1380

DB 1321 AAGATTCCTCTTTCAGCTTTTACATGTTTCCAGTAAAGATGTGCGCTCTCATGCTGTC 1380

QY 1381 TTCTGCTGCTGGGTCTATGTGTGATCATTAATTTAAAGTAAAGACGCGGTGATTTT 1440

DB 1381 TTCTGCTGCTGGGTCTATGTGTGATCATTAATTTAAAGTAAAGACGCGGTGATTTT 1440

QY 1441 TACCTCAGCATGAACATAAAGATATATACAGATGCGAGTACATACATGAAGAAATAC 1500

DB 1441 TACCTCAGCATGAACATAAAGATATATACAGATGCGAGTACATACATGAAGAAATAC 1500

QY 1501 AGAACAGTGTCTTCTAGTGGGAAACCATTTCAAGAGAAATGATCTTTGAAACAGGAAA 1560

DB 1501 AGAACAGTGTCTTCTAGTGGGAAACCATTTCAAGAGAAATGATCTTTGAAACAGGAAA 1560

QY 1561 AATGTAAGAGATCTCTCTTCACTCATTAATTTGAAGGATGGAACAGAGATAGATC 1620

DB 1561 AATGTAAGAGATCTCTCTTCACTCATTAATTTGAAGGATGGAACAGAGATAGATC 1620

QY 1621 CTGACACAGACATTAATCAGAGCTGAGTCTTTGGATGGTGGTGCACACAGTTCTACAGTTG 1680

DB 1621 CTGACACAGACATTAATCAGAGCTGAGTCTTTGGATGGTGGTGCACACAGTTCTACAGTTG 1680

QY 1681 TAGAATCTCAAGAGGGGCTTCTGCGACTCATGTCCAGAGTCTTCTGATTTGTGTAAG 1740

DB 1681 TAGAATCTCAAGAGGGGCTTCTGCGACTCATGTCCAGAGTCTTCTGATTTGTGTAAG 1740

QY 1741 GTTTTATTAATTAATTTTCAAGCAATGATATGATGGGCAAGCTTAGATTTTAAATA 1800

DB 1741 GTTTTATTAATTAATTTTCAAGCAATGATATGATGGGCAAGCTTAGATTTTAAATA 1800

QY 1801 TTGTGAAGCGCAAAAAGTGGCCCACTTAATTTAGTGTGATGCTGAACTTGCTGCTCTGA 1860

Db 1801 TTGATGAAGCGGC AAAAGTGGCCCACTAATTAGTGTGATGCTGAACCTTGATGCCCTTTCTGA 1860
Qy 1861 CAGAACAGTATCTTCAGACCACTAAACATAAAGTCTTTTGAAGAAAATGTAATGACTCTA 1920
Db 1861 CAGAACAGTATCTTCAGACCACTAAACATAAAGTCTTTTGAAGAAAATGTAATGACTCTA 1920
Qy 1921 AATCGCAAAATGAATCAGATAGATGAAGGCTTAGATGATGAAACATCAATAATATAT 1980
Db 1921 AATCGCAAAATGAATCAGATAGATGAAGGCTTAGATGATGAAACATCAATAATATAT 1980
Qy 1981 ATTCTCAATGCAAGCAGGAGCTATTGGGGAAGTCAATGATTAATAATAATTTGTGAAA 2040
Db 1981 ATTCTCAATGCAAGCAGGAGCTATTGGGGAAGTCAATGATTAATAATAATTTGTGAAA 2040
Qy 2041 CAGTGTGATAACAAATACATATAGAAATGCGCTTTCTTTAGGAGAAAAGACACTATTC 2100
Db 2041 CAGTGTGATAACAAATACATATAGAAATGCGCTTTCTTTAGGAGAAAAGACACTATTC 2100
Qy 2101 CAGTTCACCAAGGCTTACCTACCAAGTAAGTCTGAGATTACAAATCAATATCACTGCTG 2160
Db 2101 CAGTTCACCAAGGCTTACCTACCAAGTAAGTCTGAGATTACAAATCAATATCACTGCTG 2160
Qy 2161 ATATTAAACAGTCAATCTGTGGAGGGCCAGACCTTAAGCAATTTGTTAGCCTTCATCAA 2220
Db 2161 ATATTAAACAGTCAATCTGTGGAGGGCCAGACCTTAAGCAATTTGTTAGCCTTCATCAA 2220
Qy 2221 GAACAGAGGTTCAAAGGACTGTAATAAGCCAGATGTTCCAGATACAAATAGAAAGTGAAC 2280
Db 2221 GAACAGAGGTTCAAAGGACTGTAATAAGCCAGATGTTCCAGATACAAATAGAAAGTGAAC 2280
Qy 2281 CCAGCACAGCAGATACCGTTGTTCCAAATCACTGTGTCTATAGATTCTACAGCTGATCCAC 2340
Db 2281 CCAGCACAGCAGATACCGTTGTTCCAAATCACTGTGTCTATAGATTCTACAGCTGATCCAC 2340
Qy 2341 AGGTTAGCTTCAACTCTAATTACATGATATAGAAAGTAACTCTGAGGTGGATCTAGTT 2400
Db 2341 AGGTTAGCTTCAACTCTAATTACATGATATAGAAAGTAACTCTGAGGTGGATCTAGTT 2400
Qy 2401 TCGTAACTGCAAAATGAAGATTCTGTAACCTGAAAACACTTGCAAGAGGCTTGGTTTGG 2460
Db 2401 TCGTAACTGCAAAATGAAGATTCTGTAACCTGAAAACACTTGCAAGAGGCTTGGTTTGG 2460
Qy 2461 GCCAGAACAGCTACTTGGTTCCTGATTCAGAGCTCCAAACTGTATGAATGCCAAG 2520
Db 2461 GCCAGAACAGCTACTTGGTTCCTGATTCAGAGCTCCAAACTGTATGAATGCCAAG 2520
Qy 2521 TCAAAATTTACTTTTACAAACGGGACACCACTCCGAGCATGTGGGAAAAGTATTTTGG 2580
Db 2521 TCAAAATTTACTTTTACAAACGGGACACCACTCCGAGCATGTGGGAAAAGTATTTTGG 2580
Qy 2581 GTGCTGTGTPAATAGGAAGTGTAACTGCAATATCTAGAAAAGGAAAGCAAGATATGTG 2640
Db 2581 GTGCTGTGTPAATAGGAAGTGTAACTGCAATATCTAGAAAAGGAAAGCAAGATATGTG 2640
Qy 2641 TAGTCTGCTATGAACCTATTAGTAAAGCTCAGGCATTTGAAAGGATCATGATCCCACTG 2700
Db 2641 TAGTCTGCTATGAACCTATTAGTAAAGCTCAGGCATTTGAAAGGATCATGATCCCACTG 2700
Qy 2701 GTTCTAATCTTAAAGTCTAATCATCTGATGAATGACTACTGTCCAGCCTCCTCAGGAGA 2760
Db 2701 GTTCTAATCTTAAAGTCTAATCATCTGATGAATGACTACTGTCCAGCCTCCTCAGGAGA 2760
Qy 2761 ACCAAAATCCAGTATACCTTCACAGCACTTTGCCAGTCTCAGCACTTAAACAACAG 2820
Db 2761 ACCAAAATCCAGTATACCTTCACAGCACTTTGCCAGTCTCAGCACTTAAACAACAG 2820
Qy 2821 GTGTTGAAGGACTATGTTCCAAAAGAACAGAGAGATGTTTGCAGATGGTATATTCG 2880
Db 2821 GTGTTGAAGGACTATGTTCCAAAAGAACAGAGAGATGTTTGCAGATGGTATATTCG 2880
Qy 2881 CCAATGGTGAAGTTCAGATACACAAATTTATCATCTGGAAGTAAAGATGTTTCTGAAG 2940
Db 2881 CCAATGGTGAAGTTCAGATACACAAATTTATCATCTGGAAGTAAAGATGTTTCTGAAG 2940

Qy 2941 ACTTTAGTCTCTCTCACCTGATGTGCCTATGACAGTAAACACACAGTGTGATCATTTCCCAT 3000
Db 2941 ACTTTAGTCTCTCTCACCTGATGTGCCTATGACAGTAAACACACAGTGTGATCATTTCCCAT 3000
Qy 3001 CTACTACAGTGGAAAAGCCAAACAAATGACAGAGATATTACAGAAATGAGATAATTC 3060
Db 3001 CTACTACAGTGGAAAAGCCAAACAAATGACAGAGATATTACAGAAATGAGATAATTC 3060
Qy 3061 AGAGTCCCTATTCTCAGGTTCCATCAGTGGAAAATTTCTCTATGAACACAGAAATGAGG 3120
Db 3061 AGAGTCCCTATTCTCAGGTTCCATCAGTGGAAAATTTCTCTATGAACACAGAAATGAGG 3120
Qy 3121 GGTTAACCTACTCTCTGGTTCAITTTACACTAGATGATGTGTTTTGACAGAACTGAGAAC 3180
Db 3121 GGTTAACCTACTCTCTGGTTCAITTTACACTAGATGATGTGTTTTGACAGAACTGAGAAC 3180
Qy 3181 CATCTAGTCCCTACTGCTGTCTTAGTTAAACAGCAATTTACCTATTGCTAGTATTTCCAGAT 3240
Db 3181 CATCTAGTCCCTACTGCTGTCTTAGTTAAACAGCAATTTACCTATTGCTAGTATTTCCAGAT 3240
Qy 3241 ATAGGTTACTGTGCTATTAACCAAGTATGCTGCAATAAGATTAGTCTTCTACCTAATG 3300
Db 3241 ATAGGTTACTGTGCTATTAACCAAGTATGCTGCAATAAGATTAGTCTTCTACCTAATG 3300
Qy 3301 ATGAGACAGTTTGGCCCACTCTCTGGTTGCAATCTGGAGAAAAGGATCAGTGCCTGTAG 3360
Db 3301 ATGAGACAGTTTGGCCCACTCTCTGGTTGCAATCTGGAGAAAAGGATCAGTGCCTGTAG 3360
Qy 3361 TAGAAGACATCCATCTCATGAGCAGATCATTTTCTTGAAGTGAAGGCTTTTCATC 3420
Db 3361 TAGAAGACATCCATCTCATGAGCAGATCATTTTCTTGAAGTGAAGGCTTTTCATC 3420
Qy 3421 CTGTTACATTTGCTCTAAATGCTAACTACTCTGTAATGTCAAATTCATATTTATTCCT 3480
Db 3421 CTGTTACATTTGCTCTAAATGCTAACTACTCTGTAATGTCAAATTCATATTTATTCCT 3480
Qy 3481 CAGACAAATATTGGTACTTTTCAACCAATGATTCATGGCTTGGACAGGCAAGAAATTA 3540
Db 3481 CAGACAAATATTGGTACTTTTCAACCAATGATTCATGGCTTGGACAGGCAAGAAATTA 3540
Qy 3541 TTATTTCTATTGTTATGTTTGCCTAAATGAAGTACTATTCTTAAGGACATCTTCAGACTAT 3600
Db 3541 TTATTTCTATTGTTATGTTTGCCTAAATGAAGTACTATTCTTAAGGACATCTTCAGACTAT 3600
Qy 3601 TTATCACCATAATTAAGGATGCTTAAAGGAAAATACATAGAAAACCTTGGACAATATTA 3660
Db 3601 TTATCACCATAATTAAGGATGCTTAAAGGAAAATACATAGAAAACCTTGGACAATATTA 3660
Qy 3661 CCTTTACTGAGAGTTTCTCAGTAGCAAGATCAGGAGGATTCCTGTTTATTACACCTA 3720
Db 3661 CCTTTACTGAGAGTTTCTCAGTAGCAAGATCAGGAGGATTCCTGTTTATTACACCTA 3720
Qy 3721 CTTTTCAGAAAACCTTGATGATCTCTCATTAACCAAGTAACTCTTTTCTTTGTGAAATTCCTA 3780
Db 3721 CTTTTCAGAAAACCTTGATGATCTCTCATTAACCAAGTAACTCTTTTCTTTGTGAAATTCCTA 3780
Qy 3781 TCAGAGGCTTGAGATTCCTGGCAAGGTTTTCCTATGCGTTTAAATGTTGAGATTGG 3840
Db 3781 TCAGAGGCTTGAGATTCCTGGCAAGGTTTTCCTATGCGTTTAAATGTTGAGATTGG 3840
Qy 3841 GTGCAATATATAAGCATATCTGCTCTTAACCAAGCATCAGAGGCCGAAAACCTCTTT 3900
Db 3841 GTGCAATATATAAGCATATCTGCTCTTAACCAAGCATCAGAGGCCGAAAACCTCTTT 3900
Qy 3901 TTGGAGAAATAGACACACTATTATGAACCTTCTGTTGACCTTGAAATTTACCAAGTATA 3960
Db 3901 TTGGAGAAATAGACACACTATTATGAACCTTCTGTTGACCTTGAAATTTACCAAGTATA 3960
Qy 3961 CTTTGCAATATATAGATCAACTGTTGATTTCATATGAATGGCAAAAAGCTGCATAAAAA 4020
Db 3961 CTTTGCAATATATAGATCAACTGTTGATTTCATATGAATGGCAAAAAGCTGCATAAAAA 4020

QY 605 GGGACGATGTAGTAAACCTATCTGTGATCTGATTAAGTGACATGGGTAACTTAGTTCATGC 664
DB 468 GGGACGATGTAGTAAACCTATCTGTGATCTGATTAAGTGACATGGGTAACTTAGTTCATGC 527
QY 665 AACCAATAGTGAAGAAGATATTAATAAATTTATTCAGATGATTTTAAGTCTTAATGCAGA 724
DB 528 AACCAATAGTGAAGAAGATATTAATAAATTTATTCAGATGATTTTAAGTCTTAATGCAGA 587
QY 725 TTCCTTGATTTGATTTGATTTATCTTCAGTGTGATGATCTCCCTGTGTTTCTTCAACAGA 784
DB 588 TTCCTTGATTTGATTTGATTTATCTTCAGTGTGATGATCTCCCTGTGTTTCTTCAACAGA 647
QY 795 CCATGATAGTATCTGTCAGAGAACACAGATGATATCAATCTGATCTGATTTTACAAAATAG 844
DB 648 CCATGATAGTATCTGTCAGAGAACACAGATGATATCAATCTGATTTTACAAAATAG 707
QY 845 AGAAATCGGAGGAATCAAGAATTTGGGTATTAAGATAGATACAAACATTTTCAGATTCCTA 904
DB 708 AGAAATCGGAGGAATCAAGAATTTGGGTATTAAGATAGATACAAACATTTTCAGATTCCTA 767
QY 905 TAAATTACAGTGAACAGAAAATTTAAAGATATAAAGATCTTTTAAATCAGTTAGATCAAT 964
DB 768 TAAATTACAGTGAACAGAAAATTTAAAGATATAAAGATCTTTTAAATCAGTTAGATCAAT 827
QY 965 TGTGATTTTAAATGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1024
DB 828 TGTGATTTTAAATGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 887
QY 1025 AGACAAGTACAAACAGAGAGCCAGCCATGTGGATTTACTAAAGATGTTGGCTTAGTAAA 1084
DB 888 AGACAAGTACAAACAGAGAGCCAGCCATGTGGATTTACTAAAGATGTTGGCTTAGTAAA 947
QY 1085 AGAGAAGTGTAGTGGGAGGATCAATCTGCGCAGAGATGTTTAAAGAGAGGCGCAAGAC 1144
DB 948 AGAGAAGTGTAGTGGGAGGATCAATCTGCGCAGAGATGTTTAAAGAGAGGCGCAAGAC 1007
QY 1145 AAGTCTTTGACCTGCAGCCTTCGAAAATTTGAAGATTTATGCTTAAATGATCAAAATTC 1204
DB 1008 AAGTCTTTGACCTGCAGCCTTCGAAAATTTGAAGATTTATGCTTAAATGATCAAAATTC 1067
QY 1205 AAGATGAAATTTCAATTTACCTGACTTTTCCCTTCAGAGATGAGCTGTTATATA 1264
DB 1068 AAGATGAAATTTCAATTTACCTGACTTTTCCCTTCAGAGATGAGCTGTTATATA 1127
QY 1265 ACAATCTGCACAGAAGACTCAAAAAGTTTGAAGATTAATGATGATGATGATGATGATGATGAT 1324
DB 1128 ACAATCTGCACAGAAGACTCAAAAAGTTTGAAGATTAATGATGATGATGATGATGATGATGAT 1187
QY 1325 TTCCTCTTCAGCTTTACATGTTTCCAGTAAAGATGTCGCTTCATGCTGCTGCTTCC 1384
DB 1188 TTCCTCTTCAGCTTTACATGTTTCCAGTAAAGATGTCGCTTCATGCTGCTGCTTCC 1247
QY 1385 TCGCTCTGGCTCTATGCTGATGATCAATTAATGAAAGTAAAGACAGGGGTGATTTTATACC 1444
DB 1248 TCGCTCTGGCTCTATGCTGATGATCAATTAATGAAAGTAAAGACAGGGGTGATTTTATACC 1307
QY 1445 TCAGCATGAACATAAAGATATATACAGATGTCAGTACTATACATGAAGAAATACAGAA 1504
DB 1308 TCAGCATGAACATAAAGATATATACAGATGTCAGTACTATACATGAAGAAATACAGAA 1367
QY 1505 CAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1564
DB 1368 CAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1427
QY 1565 TAAAGCATCTCTCTCAGTCAATTAATTAAGAGGATGAAGACAGAAAGATAGATCTCTGA 1624
DB 1428 TAAAGCATCTCTCTCAGTCAATTAATTAAGAGGATGAAGACAGAAAGATAGATCTCTGA 1487
QY 1625 CCAGACAGTATCAGAGCTGAGTCTTTGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1684
DB 1488 CCAGACAGTATCAGAGCTGAGTCTTTGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1547
QY 1685 ATCTCAAGAGGGGCTTTCTGGCACTCATGTCCCAGAGTCTTCTGATTTGTTGAGGGTTT 1744

DB 1548 ATCTCAAGAGGGGCTTTCTGGCACTCATGTCCCAGAGTCTTCTGATTTGTTGAGGGTTT 1607
QY 1745 TATTAAATCTTTTCAAGCAATGATATGATGGGCAAGACTTAGATTAATTAATTAATTAATTAAT 1804
DB 1608 TATTAAATCTTTTCAAGCAATGATATGATGGGCAAGACTTAGATTAATTAATTAATTAATTAAT 1667
QY 1805 TGAAGCGCAAAAGTGGCCCACTAATTAGTGATGCTGAACTTGTGATGCTTCTTGACAGA 1864
DB 1668 TGAAGCGCAAAAGTGGCCCACTAATTAGTGATGCTGAACTTGTGATGCTTCTTGACAGA 1727
QY 1865 ACAGTATCTTTCAGACCACTTAACATAAAGTCTTTTGAAGAAATGTAATGATCTTAAATC 1924
DB 1728 ACAGTATCTTTCAGACCACTTAACATAAAGTCTTTTGAAGAAATGTAATGATCTTAAATC 1787
QY 1925 GCAATGATCAGATGATATGAAGGCTTAGATGATGAAACATCAATATATATATATATATAT 1984
DB 1788 GCAATGATCAGATGATATGAAGGCTTAGATGATGAAACATCAATATATATATATATATAT 1847
QY 1985 CAATGCAAGCAGAGAGCTTATTTGGGAAAAGTCAATGATTAATTAATTAATTAATTAATTAAT 2044
DB 1848 CAATGCAAGCAGAGAGCTTATTTGGGAAAAGTCAATGATTAATTAATTAATTAATTAATTAAT 1907
QY 2045 TGATAAACAATAATACATAGAAAATTTGGGCTTTCTTTAGAGAAAAGCACTATTTCCAGT 2104
DB 1908 TGATAAACAATAATACATAGAAAATTTGGGCTTTCTTTAGAGAAAAGCACTATTTCCAGT 1967
QY 2105 TCAACAAGGGTTCCTTACCAGTAAGTCTGAGATTAACAATCAATTAATTAATTAATTAATTAAT 2164
DB 1968 TCAACAAGGGTTCCTTACCAGTAAGTCTGAGATTAACAATCAATTAATTAATTAATTAATTAAT 2027
QY 2165 TAAAGTCAATCTGTTGGAGGGCCAGACCTTAAGCAATTTGTTTACCTTCCATCAAGAAC 2224
DB 2028 TAAAGTCAATCTGTTGGAGGGCCAGACCTTAAGCAATTTGTTTACCTTCCATCAAGAAC 2087
QY 2225 AAGGAGTTCAGAGGACCTGAATAAGCCAGATGTTCCAGATACAAATAGAAAAGTGAACCCAG 2284
DB 2088 AAGGAGTTCAGAGGACCTGAATAAGCCAGATGTTCCAGATACAAATAGAAAAGTGAACCCAG 2147
QY 2285 CACAGCAGATACCGTTTCCAAATCACTTGTCTATAGATTTCTACAGCTGATCCACAGT 2344
DB 2148 CACAGCAGATACCGTTTCCAAATCACTTGTCTATAGATTTCTACAGCTGATCCACAGT 2207
QY 2345 TAGCTTCAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2404
DB 2208 TAGCTTCAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2267
QY 2405 AACTGCAAAATGAAGATTTCTGTACCTGAAAACACTTGCAAAAGAGCTTGGTTTGGGCA 2464
DB 2268 AACTGCAAAATGAAGATTTCTGTACCTGAAAACACTTGCAAAAGAGCTTGGTTTGGGCA 2327
QY 2465 GAAACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2524
DB 2328 GAAACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2387
QY 2525 ATTTACTTTTACCAACGGCGACACCATTTCCGAGCATGTGGGAAAGTATTTTGTGGTGT 2584
DB 2388 ATTTACTTTTACCAACGGCGACACCATTTCCGAGCATGTGGGAAAGTATTTTGTGGTGT 2447
QY 2585 CTGTTCTTAATAGAAAGTGAATGCAATATCTAGAAAAGGAAAGCAAGATGATGTTAGT 2644
DB 2448 CTGTTCTTAATAGAAAGTGAATGCAATATCTAGAAAAGGAAAGCAAGATGATGTTAGT 2507
QY 2645 CTGCTATGAATCTTATAGTAAAGCTCAGGCAATTTGAAAGGATGATGAGTCCAACTGGTTC 2704
DB 2508 CTGCTATGAATCTTATAGTAAAGCTCAGGCAATTTGAAAGGATGATGAGTCCAACTGGTTC 2567
QY 2705 TAACTTAAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2764
DB 2568 TAACTTAAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2627
QY 2765 AACATCAGTATACCTTCCAGCACTTTGCGAGTCTCAGCACTTAAACCAACAGGTTGT 2824

Db 2628 AACATCCAGTATACCTTACCAGCAACTTTGCCAGTCTCAGCACTTAAACAAACAGGTGT 2687
Qy 2825 TGAAGGACTATGTTCCAAAGAACAGAGAGATGTTTGCAGATGATATATTTGCCCAA 2684
Db 2688 TGAAGGACTATGTTCCAAAGAACAGAGAGATGTTTGCAGATGATATATTTGCCCAA 2747
Qy 2885 TGGTGAAGTTCAGATACAAACAAATATATCATCTGGAAGTAAAGATGTTCTGAGACTT 2944
Db 2748 TGGTGAAGTTCAGATACAAACAAATATATCATCTGGAAGTAAAGATGTTCTGAGACTT 2807
Qy 2945 TAGTCCTCTCTCAGCTGATGTCCTATGACAGTAAACACAGTGGATCATTTCCCATTTCTAC 3004
Db 2808 TAGTCCTCTCTCAGCTGATGTCCTATGACAGTAAACACAGTGGATCATTTCCCATTTCTAC 2867
Qy 3005 TACAGTGGAAAGCCAAACATGACAGAGAGATATTTACAAGAAATCAGATAATTTACAG 3064
Db 2868 TACAGTGGAAAGCCAAACATGACAGAGAGATATTTACAAGAAATCAGATAATTTACAG 2927
Qy 3065 TCCTATTTCTCAGGTTTCATGAGTGGAAATTTGTCTATGAAACACAGGAAATGAGGGTT 3124
Db 2928 TCCTATTTCTCAGGTTTCATGAGTGGAAATTTGTCTATGAAACACAGGAAATGAGGGTT 2987
Qy 3125 ACCTACTTCTGGTTCATTTACACTAGATGATGTTTTCAGAAACTGAAGAACCATC 3184
Db 2988 ACCTACTTCTGGTTCATTTACACTAGATGATGTTTTCAGAAACTGAAGAACCATC 3047
Qy 3185 TAGTCCTACTGGTCTCTTAGTTAAACAGCAATTTACCTATTGCTAGTATTTACAGATTATAG 3244
Db 3048 TAGTCCTACTGGTCTCTTAGTTAAACAGCAATTTACCTATTGCTAGTATTTACAGATTATAG 3107
Qy 3245 GTTACTGTGTATATTAACAGTATGTCGCAATAAGATAGTCTTCTACTTAATGATGA 3304
Db 3108 GTTACTGTGTATATTAACAGTATGTCGCAATAAGATAGTCTTCTACTTAATGATGA 3167
Qy 3305 GGACAGTTTGGCCCACTCTCTGGTTCATCTGGAGAAAGGGATCAGTGCCTGTAGTAGA 3364
Db 3168 GGACAGTTTGGCCCACTCTCTGGTTCATCTGGAGAAAGGGATCAGTGCCTGTAGTAGA 3227
Qy 3365 AGAACATCCATCTCATGAGCAGATCTTTGCTTCTTGAAGTGAAGGCTTTCATCCTGT 3424
Db 3228 AGAACATCCATCTCATGAGCAGATCTTTGCTTCTTGAAGTGAAGGCTTTCATCCTGT 3287
Qy 3425 TACATTTGCTCTAAATGCTAATCTPACTCGTGAATGTCAAATTCATATTTTATTCCTCAGA 3484
Db 3288 TACATTTGCTCTAAATGCTAATCTPACTCGTGAATGTCAAATTCATATTTTATTCCTCAGA 3347
Qy 3485 CAAATATGGTACTTTTCAACCAATGGAATGCAATGCTGGACAGGACAGCAATATTAT 3544
Db 3348 CAAATATGGTACTTTTCAACCAATGGAATGCAATGCTGGACAGGACAGCAATATTAT 3407
Qy 3545 TCTATTGTTATGTTTGCCTAATGAAGATCTATTCTCTAAGGACATCTTCAGACTATTAT 3604
Db 3408 TCTATTGTTATGTTTGCCTAATGAAGATCTATTCTCTAAGGACATCTTCAGACTATTAT 3467
Qy 3605 CACCATATTAAGGATGCTCTTAAAGGAAATACATAGAAAACCTTGGACAAATATTACCTT 3664
Db 3468 CACCATATTAAGGATGCTCTTAAAGGAAATACATAGAAAACCTTGGACAAATATTACCTT 3527
Qy 3665 TACTGAGATTTTCTAGTAGCAAGATCAGGAGGATTCCTGTTTATTTACACTACTTT 3724
Db 3528 TACTGAGATTTTCTAGTAGCAAGATCAGGAGGATTCCTGTTTATTTACACTACTTT 3587
Qy 3725 TCAGAAACTTGATGATCTCTCAATACCAAGTAATPCTTTTCTTTGTGGAATTCCTATCCA 3784
Db 3588 TCAGAAACTTGATGATCTCTCAATACCAAGTAATPCTTTTCTTTGTGGAATTCCTATCCA 3647
Qy 3785 GAAGCTTGAGATTCCTCGGGCAAGGTTTTTCTATGCGTTTTTAAATGTTGAGATGGGTGC 3844
Db 3648 GAAGCTTGAGATTCCTCGGGCAAGGTTTTTCTATGCGTTTTTAAATGTTGAGATGGGTGC 3707
Qy 3845 AGAATATAAGCATATCTCTCTCTCTAAACAGCATCAGAGCGCGAAACCTCTTTTGG 3904
Db 3708 AGAATATAAGCATATCTCTCTCTCTCTAAACAGCATCAGAGCGCGAAACCTCTTTTGG 3767

Qy 3905 AGAATATAGGACACACTATTATGAACCTTACTTGTGTGACCTTGGAAATTACCAGTATACCTT 3964
Db 3768 AGAATATAGGACACACTATTATGAACCTTACTTGTGTGACCTTGGAAATTACCAGTATACCTT 3827
Qy 3965 GCATATAATAGATCAACTGTTGATTCATATGAAATGGGAAAGCTGCATAAAAATACC 4024
Db 3828 GCATATAATAGATCAACTGTTGATTCATATGAAATGGGAAAGCTGCATAAAAATACC 3887
Qy 4025 ACGGAAAAAGTACAGTGTATGAAGTACTAAATCTTCCAAATGAGCATGTCATTAG 4084
Db 3888 ACGGAAAAAGTACAGTGTATGAAGTACTAAATCTTCCAAATGAGCATGTCATTAG 3947
Qy 4085 CATTTGAGCAAGTTTCAGTACAGAAAGCAGATTCTCATCTAGTCTCTATACAGAAATGATGG 4144
Db 3948 CATTTGAGCAAGTTTCAGTACAGAAAGCAGATTCTCATCTAGTCTCTATACAGAAATGATGG 4007
Qy 4145 AATTTATGAAACACAGGCCAACAGTGCCTGCTGCTCCTAGAAAAAGTGCAGTGCAG 4204
Db 4008 AATTTATGAAACACAGGCCAACAGTGCCTGCTGCTCCTAGAAAAAGTGCAGTGCAG 4067
Qy 4205 TTTTGTGGTATTCAATGGAGCTCTAAAAACATCTTTCCAGGATTTCTTGTAAAGTCCAGCAT 4264
Db 4068 TTTTGTGGTATTCAATGGAGCTCTAAAAACATCTTTCCAGGATTTCTTGTAAAGTCCAGCAT 4127
Qy 4265 AGTTGAAGTGGCTTAAATGGTACAAATAACTCCAGAGACCATGAAATGGCTTGGCGTAGC 4324
Db 4128 AGTTGAAGTGGCTTAAATGGTACAAATAACTCCAGAGACCATGAAATGGCTTGGCGTAGC 4187
Qy 4325 TTTTACGAGAACAGAAAGACTTTTAAATACATGTCGGAAAGTTGATGCAGTAGACTGAG 4384
Db 4188 TTTTACGAGAACAGAAAGACTTTTAAATACATGTCGGAAAGTTGATGCAGTAGACTGAG 4247
Qy 4385 AGAATACGTGGATATCTGCTGGTGTAGTCTGGAAGAAAAAGGAAACAAAGGAGTTTATCAG 4444
Db 4248 AGAATACGTGGATATCTGCTGGTGTAGTCTGGAAGAAAAAGGAAACAAAGGAGTTTATCAG 4307
Qy 4445 TTCAGTGGATGGAATATCATTTACAGGATTTCCAAAGTGAAGAAATAAACTGGAGCAGA 4504
Db 4308 TTCAGTGGATGGAATATCATTTACAGGATTTCCAAAGTGAAGAAATAAACTGGAGCAGA 4367
Qy 4505 TTTTGAACCCGATGAGAGATTTAAATGTACCCAGGTTGTCTACTTTCTTAAAGGACCA 4564
Db 4368 TTTTGAACCCGATGAGAGATTTAAATGTACCCAGGTTGTCTACTTTCTTAAAGGACCA 4427
Qy 4565 GGATTTATCTATTTTATCAACTTCTTATCAGTTTGCAGAAAGAAATAGCCATGGCTTTGAG 4624
Db 4428 GGATTTATCTATTTTATCAACTTCTTATCAGTTTGCAGAAAGAAATAGCCATGGCTTTGAG 4487
Qy 4625 TGTGCGCTGTCCCTCAGCTGAAACCTCTAAGAAAGTAAATGGGATGAATAATTTGGACT 4684
Db 4488 TGTGCGCTGTCCCTCAGCTGAAACCTCTAAGAAAGTAAATGGGATGAATAATTTGGACT 4547
Qy 4685 CAGAGTTTCCATTTGACACTGATATGTTTGAATTTTCAGGACGAGATCTGAAGGCCAACTTCT 4744
Db 4548 CAGAGTTTCCATTTGACACTGATATGTTTGAATTTTCAGGACGAGATCTGAAGGCCAACTTCT 4607
Qy 4745 GCCTCAGCATTTATCTAAATGATCTTGATAGTCTCTGATACCTGTGATCCATGGTGGAC 4804
Db 4608 GCCTCAGCATTTATCTAAATGATCTTGATAGTCTCTGATACCTGTGATCCATGGTGGAC 4667
Qy 4805 CTCCTAATCTAGTTTACCAATAGAAATAGATTAGTGTGTTTTTTCATTATAGAACACTCTTTT 4864
Db 4668 CTCCTAATCTAGTTTACCAATAGAAATAGATTAGTGTGTTTTTTCATTATAGAACACTCTTTT 4727
Qy 4865 TTAGTGAAGATGTCGCAATATACATATGCAACCTTAATTTGTTTAAACTTAACCTCCAGC 4924
Db 4728 TTAGTGAAGATGTCGCAATATACATATGCAACCTTAATTTGTTTAAACTTAACCTCCAGC 4787
Qy 4925 ACTTAAAGCTGAATGCCCAACACACTTAAAGTATATAATATGCTGATTTTGTGAAACACAT 4984
Db 4788 ACTTAAAGCTGAATGCCCAACACACTTAAAGTATATAATATGCTGATTTTGTGAAACACAT 4847

QY	4985	AAGCTTTGCTCTTTAGGAGGAGTATCTTTTCAAAATCATTAGACAAATATTAAATATC	5044
DB	4848	AAGCTTTGCTCTTTAGGAGGAGTATCTTTTCAAAATCATTAGACAAATATTAAATATC	4907
QY	5045	TAAAAATTAAGAGATCCATCTTTCTAGCTTTTACAAATTAATTAAGTACTAAAAAGA	5104
DB	4908	TAAAAATTAAGAGATCCATCTTTCTAGCTTTTACAAATTAATTAAGTACTAAAAAGA	4962
QY	5105	CAAGATTTCTTTTAAAGAAATTTATAGCAATTTACTGTTGTTTAAATGTCAGCCAAAG	5164
DB	4963	CAAGATTTCTTTTAAAGAAATTTATAGCAATTTACTGTTGTTTAAATGTCAGCCAAAG	5022
QY	5165	TATCTGCACTTAGGTATACCTCTTTATGCGCAATTAATGATTTTAAATGAAGCTCTTTTCAG	5224
DB	5023	TATCTGCACTTAGGTATACCTCTTTATGCGCAATTAATGATTTTAAATGAAGCTCTTTTCAG	5082
QY	5225	ATGTAACCTTATGAAGGAATATCTGCTTTGTTGTTATATGCCAGTTAGAACTGGTTTCT	5284
DB	5083	ATGTAACCTTATGAAGGAATATCTGCTTTGTTGTTATATGCCAGTTAGAACTGGTTTCT	5142
QY	5285	AAAGCTCTCAAAATGTTATTTTCGTGGCAGCAAAACCAAGTTTGAAGCTCTTACACTTATA	5344
DB	5143	AAAGCTCTCAAAATGTTATTTTCGTGGCAGCAAAACCAAGTTTGAAGCTCTTACACTTATA	5202
QY	5345	ATTTCTTTGAATAAACTGATAACTTATTTGTTATTAATTTGGAGTGAGACCTACCTCCATAA	5404
DB	5203	ATTTCTTTGAATAAACTGATAACTTATTTGTTATTAATTTGGAGTGAGACCTACCTCCATAA	5262
QY	5405	TTAGATAAACTCTTTTGGATTAATATCAGAAATTTTGGCTTTTCTCTCAAAATTTATA	5464
DB	5263	TTAGATAAACTCTTTTGGATTAATATCAGAAATTTTGGCTTTTCTCTCAAAATTTATA	5322
QY	5465	CATATGTTATGTTATATATATCCACATATATAGTTTTCCTGATTAAATGGATTTAAAT	5524
DB	5323	CATATGTTATGTTATATATATCCACATATATAGTTTTCCTGATTAAATGGATTTAAAT	5382
QY	5525	AA 5526	
DB	5383	AA 5384	
RESULT 3			
ID	AAC98041 standard; cDNA; 1572 BP.		
XX	AAC98041;		
AC	AAC98041;		
XX	09-MAR-2001 (first entry)		
DE	Human colon cancer antigen nucleotide sequence SEQ ID NO:51.		
XX	Human; colon cancer; colon cancer antigen; diagnosis; detection;		
KW	identification; cytostatic; cardioactive; neuroprotective; vulnery;		
KW	immunomodulatory; muscular; gynaecological; gastrointestinal;		
KW	nephrotropic; anti-infective; antibacterial; gene therapy; wound;		
KW	neural disorder; immune system disorder; muscular disorder;		
KW	reproductive disorder; gastrointestinal disorder; renal disorder;		
OS	infectious disease; cardiovascular disorder; ss.		
XX	Homo sapiens.		
OS	WO20005351-A1.		
PN	21-SEP-2000.		
XX	08-MAR-2000; 2000WO-US005883.		
XX	12-MAR-1999; 99US-0124270P.		
PR	(HUMA-) HUMAN GENOME SCI INC.		
XX	Rosen CA, Ruben SM;		
XX			
PI			
XX			

DR	WPI: 2000-587534/55.		
DR	P-PSDB; AAB53284.		
XX	Colon cancer associated gene sequences, referred to as colon cancer		
PT	antigens, useful for the treatment, prevention, and diagnosis of colon		
PT	disorders such as colon cancer.		
XX	Claim 1; Page 504-505; 2104pp; English.		
XX	AAC97991 to AAC98763 encode the human colon cancer associated proteins,		
CC	called human colon cancer antigens, given in AAB53234 to AAB54006. The		
CC	human colon cancer antigens can have cytostatic, cardioactive, muscular;		
CC	neuroprotective, immunomodulatory, gynaecological, gastrointestinal,		
CC	vulnery, nephrotropic, anti-infective and antibacterial activities, and		
CC	can be used in gene therapy. The colon cancer antigen polynucleotides,		
CC	proteins and antibodies to the proteins are useful for the prevention,		
CC	treatment and diagnosis of colon disorders, such as colon cancer. The		
CC	polynucleotides may be used in diagnostics and research, such as for		
CC	chromosome identification, and as hybridisation probes. The proteins may		
CC	also be used to prevent diseases such as neural disorders, immune system		
CC	disorders, muscular disorders, reproductive disorders, gastrointestinal		
CC	disorders, wounds, renal disorders, infectious diseases, and		
CC	cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent		
CC	sequences used in the exemplification of the present invention		
XX	Sequence 1572 BP; 537 A; 233 C; 231 G; 562 T; 0 U; 9 Other;		
SQ	Query Match 22.8%; Score 1514; DB 3; Length 1572;		
	Best Local Similarity 99.3%; Pred. No. 5.9e-287;		
	Matches 1522; Conservative 7; Mismatches 3; Indels 1; Gaps 1;		
QY	4865	TTAGTGAAGAATGTGCGATATTACATATTGCAACCTAATTTGTAAACCTAACCTCCAGC	4924
DB	1	TNAGTGAAGAATGTGCGATATTACATATTGCAACCTAATTTGTAAACCTAACCTCCAGC	60
QY	4925	ACTAAGCTGAAATGCCACAACTAAAAGTATAAATATGCTGATTTTGAACACAT	4984
DB	61	ACTAAGCTGAAATGCCACAACTAAAAGTATAAATATGCTGATTTTGAACACAT	120
QY	4985	AAGCTTTGCTCTTTAGGAGGAGTATCTTTTCAAAATCATTAGACAAATTTAAATATC	5044
DB	121	AAGCTTTGCTCTTTAGGAGGAGTATCTTTTCAAAATCATTAGACAAATTTAAATATC	180
QY	5045	TAAAAATTAAGAGATCCATCTTTCTAGCTTTTACAAATTAATTAAGTACTAAAAAGA	5104
DB	181	TAAAAATTAAGAGATCCATCTTTCTAGCTTTTACAAATTAATTAAGTACTAAAAAGA	240
QY	5105	CAAGATTTCTTTTAAAGAAATTTATAGCAATTTACTGTTGTTTAAATGTCAGCCAAAG	5164
DB	241	CAAGATTTCTTTTAAAGAAATTTATAGCAATTTACTGTTGTTTAAATGTCAGCCAAAG	300
QY	5165	TATCTGCACTTAGGTATACCTCTTTATGCGCAATTAATGATTTTAAATGTCAGCCAAAG	5224
DB	301	TATCTGCACTTAGGTATACCTCTTTATGCGCAATTAATGATTTTAAATGTCAGCCAAAG	360
QY	5225	ATGTAACCTTATGAAGGAATATCTGCTTTGTTGTTATATGCCAGTTAGAACTGGTTTCT	5284
DB	361	ATGTAACCTTATGAAGGAATATCTGCTTTGTTGTTATATGCCAGTTAGAACTGGTTTCT	420
QY	5285	AAAGCTCTCAAAATGTTATTTTCGTGGCAGCAAAACCAAGTTTGAAGCTCTTACACTATA	5344
DB	421	AAAGCTCTCAAAATGTTATTTTCGTGGCAGCAAAACCAAGTTTGAAGCTCTTACACTATA	480
QY	5345	ATTTCTTTGAATAAACTGATAACTTATTTGTTATTAATTTGGAGTGAGACCTACCTCCATAA	5404
DB	481	ATTTCTTTGAATAAACTGATAACTTATTTGTTATTAATTTGGAGTGAGACCTACCTCCATAA	540
QY	5405	TTAGATAAACTCTTTTGGATTAATATCAGAAATTTTGGCTTTTCTCTCAAAATTTATA	5464
DB	541	TTAGATAAACTCTTTTGGATTAATATCAGAAATTTTGGCTTTTCTCTCAAAATTTATA	600
QY	5465	CATATGTTATGTTATATATATCCACATATATAGTTTTCCTGATTAAATGGATTTAAAT	5524

QY	3393	TTGTTCTTTGAAGTGAAGGCTTTTCATCTCTGTTACATTTGTCCTTAATGCTATCTACTC	3452	PN	WO200021991-A1.
DB	301	TTGTTCTTTGAAGTGAAGGCTTTTCATCTCTGTTACATTTGTCCTTAATGCTATCTACTC	360	FD	20-APR-2000.
QY	3453	GTGAATGTCATATTCATATTTATTTCTCTCAGACAAATATTTGGTACTTTTCAACCAATGGA	3512	PF	15-OCT-1999; 99WO-US024206.
DB	361	GTGAATGTCATATTCATATTTATTTCTCTCAGACAAATATTTGGTACTTTTCAACCAATGGA	419	XX	15-OCT-1998; 98US-0104436P.
QY	3513	TTGCATGCTTTGGACAGCGCAGAAATATTTATTTCTATTTGTTATGTTGCGCAATGAGAT	3572	XX	(GENY) GENETICS INST INC.
DB	420	TTGCA--GGCTGGACAGCGGAGAAATATTTA-TCTATTTGTTATGTTGCGCAATGAGAT	476	XX	Jacobs K, McCoey JM, Lavallie ER, Collins-Racie LA, Evans C;
QY	3573	ACTATTTCTAAGGACATCTTCAGACTATTTATCCACATATATAGGATGCTCTAAAGGA	3632	PI	Merberg D, Treacy M, Bowman MR;
DB	477	ACTATTTCTAAGGACATCTTCAGACTATTTATCCACATATATAGGATGCTCTAAAGGA	536	XX	WPI; 2000-317938/27.
QY	3633	AAATACATAGAAACTTTGGACAAATATTTACTCTTCTCAGTAGCAAGGAT	3692	XX	Isolated polynucleotides, and encoded proteins, comprising secreted
DB	537	AAATACATAGAAACTTTGGACAAATATTTACTCTTCTCAGTAGCAAGGAT	596	PT	expressed sequence tags (sESTs), useful for treating various disorders
QY	3693	CACGAGGATTCCTGTTTATTAACCTTCTTTTTCAGAACTTGTATGATCTCTCATACCA	3752	PT	such as autoimmune, infectious, and central nervous system disorders.
DB	597	CACGAGGATTCCTGTTTATTAACCTTCTTTTTCAGAACTTGTATGATCTCTCATACCA	656	XX	Claim 1; Page 443; 803pp; English.
QY	3753	AGTAATCCTTTTCTTTGGAATTCCTTATCCAGAGCTTTGAGATTCCTCGGCAAGGTT	3812	CC	AAA43426 to AAA45925 represent specifically claimed expressed
DB	657	AGTAATCCTTTTCTTTGGAATTCCTTATCCAGAGCTTTGAGATTCCTCGGCAAGGTT	716	CC	sequence tags (sESTs), isolated from human, mouse, chicken and rat tissue
QY	3813	TTTCTATGCTTTAATGTTGAGATTTGGTGAGCAATATTAAGCATATCTGTCCTCTA	3872	CC	sources. The sESTs can have a range of activities depending on the
DB	717	TTTCTATGCTTTAATGTTGAGATTTGGTGAGCAATATTAAGCATATCTGTCCTCTA	776	CC	tissues they were isolated from. The activities include: chemotactic;
QY	3873	ACAGCATCAGAGCGGAAACCTCTTTTGGAGAAATAGACACACTATTTAAGACTTA	3932	CC	proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic;
DB	777	ACAGCATCAGAGCGGAAACCTCTTTTGGAGAAATAGACACACTATTTAAGACTTA	836	CC	haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial;
QY	3933	CTTGTGACCTTCGAAATACAGTATACCTTGCA-TAATATAGATCAACTGTG-ATTC	3990	CC	antifungal; antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;
DB	837	CTTGTGACCTTCGAAATACAGTATACCTTGCA-TAATATAGATCAACTGTG-ATTC	896	CC	osteopathic; neuroprotective; nootropic; antiparkinsonian; antipsoriatic;
QY	3991	ATATGGAATCGGAAAGCTGCTTAAATATACACGAAAA	4032	CC	cerebroprotective; anticonvulsant; and antidepressant. The sESTs can be
DB	897	NTNTGGAATCGGAAAGCTGCTTAAATATACACGAAAA	938	CC	used for gene therapy and in vaccines. The sESTs are useful as probes for
RESULT 5					the identification and isolation of full-length cDNAs and genomic DNA
AAA44354					molecules which correspond to the sESTs. Proteins encoded by the sESTs
ID AAA44354 standard; cDNA; 693 BP.					are useful in assays for determining biological activity and raising
AC AAA44354;					antibodies. They may be useful for treatment of autoimmune disorders
AC AAA44354;					(multiple sclerosis, insulin dependent diabetes), allergic conditions
DT 21-AUG-2000 (first entry)					(asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
XX Human secreted expressed sequence tag SEQ ID NO:929.					osteoporosis, osteoarthritis, central nervous system disorders
XX Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;					(Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
XX expressed sequence tag; EST; probe; chemotactic; proliferative;					disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;					disease), tumours, bacterial, fungal or viral infections, depression and
XX thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;					psoriasis). AAA45926 to AAA45931 represent linker variants which are given
XX antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;					in the exemplification of the present invention
XX antitumor; osteopathic; neuroprotective; nootropic; antipsoriatic;					
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine;					
XX autoimmune disorder; multiple sclerosis; allergic condition;					
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;					
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;					
XX central nervous system disorder; Alzheimer's disease; stroke;					
XX Parkinson's disease; Huntington's disease; coagulation disorder;					
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour;					
XX infection; depression; psoriasis; ss.					
OS Homo sapiens.					
XX					

Query Match 10.0%; Score 665.4; DB 3; Length 693;

Best Local Similarity 99.1%; Pred. No. 1.1e-120;

Matches 669; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 853 GAGGATCAAGAAATGGGTATAAAGTAGATACACACTTTCAGATTCCTATATTACA 912

DB 14 GAGGCTAAAGAAATGGGTATAAAGTAGATACACACTTTCAGATTCCTATATTACA 73

QY 913 GTGGAACAGAAAATTTAAAGATAAAAGATCTTTTAATCAGTTAGATCAATTTGTGATT 972

DB 74 GTGGAACAGAAAATTTAAAGATAAAAGATCTTTTAATCAGTTAGATCAATTTGTGATT 133

QY 973 TTAACTGTCATCTGCTTTGACTCGACAAAGTTCAAAATGTTTCATGCCAAGCAAGC 1032

DB 134 TTAACTGTCATCTGCTTTGACTCGACAAAGTTCAAAATGTTTCATGCCAAGCAAGC 193

QY 1033 TACAACACAAAGCCAGCCATCTGATTACTAAAAGATGTTGGCTTAGTAAAAGAGGAAG 1092

DB 194 TACAACACAAAGCCAGCCATCTGATTACTAAAAGATGTTGGCTTAGTAAAAGAGGAAG 253

QY 1093 TAGATGTGTCAGTCAATCTGCGGAGCAATGTTTAAAGACAGCGGCAAGCAAGTGT 1152

DB 254 TAGATGTGTCAGTCAATCTGCGGAGCAATGTTTAAAGACAGCGGCAAGCAAGTGT 313

QY 1153 TGACCTGCAGCCTTCGAAAAATGAAGATTTATGCTTAAATGATTCAAAATCAAGAGATG 1212

DB 314 TGACCTGCAGCCTTCGAAAAATGAAGATTTATGCTTAAATGATTCAAAATCAAGAGATG 373

QY 1213 AAAATTTCAATACCTGACTTTTCCTTTCCAGGAAGATAGACTGTATATAAAACAATCTG 1272
Db |||||
QY 374 ABAATTTCAATACCTGACTTTTCCTTTCCAGGAAGATAGACTGTATATAAAACAATCTG 433
Db |||||
QY 1273 CACAAGAAGACTCAAAAGTTTACACCTTAAGGATATGATGTAATCCAAAGATTCCTCTT 1332
Db |||||
QY 434 CACAAGAAGACTCAAAAGTTTACACCTTAAGGATATGATGTAATCCAAAGATTCCTCTT 493
Db |||||
QY 1333 CAGCTTTACATGTTTCCAGTAAGATGTCGCTCCTCATTTGCTCTGCTTCCTCGCTCTG 1392
Db |||||
QY 494 CAGCTTTACATGTTTCCAGTAAGATGTCGCTCCTCATTTGCTCTGCTTCCTCGCTCTG 553
Db |||||
QY 1393 GGCTATGTGTGATCAATTAATGAAGTAAGACACGGGGTGATTTTTTACCTCAGCATG 1452
Db |||||
QY 554 GGCTATGTGTGATCAATTAATGAAGTAAGACACGGGGTGATTTTTTACCTCAGCATG 613
Db |||||
QY 1453 AACATAAGATAATATACAGATGCGAGTGACTATACATGAAGAAATACAGACAGGTGTG 1512
Db |||||
QY 614 AACATAAGATAATATACAGATGCGAGTGACTATACATGAAGAAATACAGACAGGTGTG 673
Db |||||
QY 1513 TTCTAGTGGGGAAC 1527
Db |||||
QY 674 TTCTAGATGGGGAAC 688
Db |||||

RESULT 6
ABQ57538
ID ABQ57538 standard; cDNA; 701 BP.
AC ABQ57538;
XX
XX
DT 02-AUG-2002 (first entry)
XX
XX
DE Human colon cancer related nucleotide sequence SEQ ID NO:1233.
KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;
KW genetic analysis; diagnostic; therapeutic; antisense therapy; gene; ss.
OS Homo sapiens.
FN WO200229086-A2.
XX
XX
PD 11-APR-2002.
XX
XX
PF 02-OCT-2001; 2001WO-US030732.
XX
XX
PR 02-OCT-2000; 2000US-0237271P.
XX
XX
PA (FARB) BAYER CORP.
XX
XX
PI Burges C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
PI Thiagalingam A, Lewis ME;
XX
XX
DR WPI; 2002-426115/45.
XX
XX

New isolated nucleic acid that is differentially expressed in cancer tissues useful for determining the presence of colon cancer in a cell or tissue type, and in antisense therapy.

Claim 1; Fig 1; 796pp; English.

ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be used in antisense therapy. An antibody immunoreactive with a polypeptide encoded by (I) is useful for detecting cancer in a patient sample, and for detecting the presence or absence of a polynucleotide encoded by a nucleic acid which hybridizes to (I) in a cell. A probe/primer derived from (I) can be used for determining the presence of a nucleic acid which hybridizes to (I), and for determining the phenotype of cells in a sample of cells from a patient. (I) is useful for determining the presence of colon cancer in a cell or tissue type, for determining the presence or

CC state of other type of cancer, in antisense therapy, to generate
CC macroarrays on a solid surface, to identify a chromosome on which the
CC corresponding gene resides, and in tissue profiling, forensics, genetic
CC analysis, mapping and diagnostic applications. (I) can be used to raise
CC antibodies, and to screen for peptide analogues and antagonists
XX
SQ Sequence 701 BP; 220 A; 134 C; 142 G; 197 T; 0 U; 8 Other;

Query Match 8.7%; Score 580.2; DB 6; Length 701;
Best Local Similarity 99.0%; Pred. No. 5.4e-104;
Matches 604; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 2736 ACTACTGTCAGCCCTCCTCAGGAGAACCAACATCCAGTATACCTTCCACGACCACTTTG 2795
Db 1 ACTACTGTCAGCCCTCCTCAGGAGAACCAACATCCAGTATACCTTCCACGACCACTTTG 60
QY 2796 CCAGTCTCAGCACTTAAACCAACAGGTGTTGAAGACTATGTTCCAAAGACAGAGAGA 2855
Db 61 CCAGTCTCAGCACTTAAACCAACAGGTGTTGAAGACTATGTTCCAAAGACAGAGAGA 120
QY 2856 GTATGTTTGCAGATGGTATATTTGCCAATGGTGAAGTTGCAGATACAAACAAATATCA 2915
Db 121 GTATGTTTGCAGATGGTATATTTGCCAATGGTGAAGTTGCAGATACAAACAAATATCA 180
QY 2916 TCTGGAAGTAAAGATGTTCTGAAGACTTTAGTCTCTCTCAGCTGATGCTCCTATGACA 2975
Db 181 TCTGGAAGTAAAGATGTTCTGAAGACTTTAGTCTCTCTCAGCTGATGCTCCTATGACA 240
QY 2976 GTAAACACAGTGGATCATTTCCCATTTCTACTACAGTGGAAAGCCAAACAAATGAGACAGGA 3035
Db 241 GTAAACACAGTGGATCATTTCCCATTTCTACTACAGTGGAAAGCCAAACAAATGAGACAGGA 300
QY 3036 GATATTACAGAAATGAGATTAATTCAGAGTCTTATTTCTCAGTTCCATCAGTGGAAAAA 3095
Db 301 GATATTACAGAAATGAGATTAATTCAGAGTCTTATTTCTCAGTTCCATCAGTGGAAAAA 360
QY 3096 TTGCTCTATCAACACAGGAAATGAGGGGTTACTACTCTGTTTCACTTACACTAGATCAT 3155
Db 361 TTGCTCTATCAACACAGGAAATGAGGGGTTACTACTCTGTTTCACTTACACTAGATCAT 420
QY 3156 GATGTTTTCAGAACTGAAGAACCATCTAGTCTCTACTGTTCTTGTCTTGTAGTTAAACAGCAAT 3215
Db 421 GATGTTTTCAGAACTGAAGAACCATCTAGTCTCTACTGTTCTTGTCTTGTAGTTAAACAGCAAT 480
QY 3216 TTACCTATTGCTAGTATTTCAGATTATAGTTTACTGTTGATATTAAAC- AAGATGTTCTG 3274
Db 481 TTACCTATTGCTAGTATTTCAGATTATAGTTTACTGTTGATATTAAACAGATGTTCTG 540
QY 3275 CAATAAGATTAGTCTTCTACCTAATGATG-AGGACAGTTTCCCTCCACTTCTGTTGAT 3333
Db 541 CAATAAGATTAGTCTTCTACCTAATGATGAGGACAGTTTCCCTCCACTTCTGTTGAT 600
QY 3334 CTGGAGAAAA 3343
Db 601 CTGGAGAAAA 610

RESULT 7
AAZ50068

ID AAZ50068 standard; cDNA; 4839 BP.

XX AAZ50068;

XX AAZ50068;

DT 04-MAY-2000 (first entry)

XX Human Smad Anchor for Receptor Activation protein-1 encoding cDNA.

DE Smad Anchor for Receptor Activation protein; hSAR1; human;
KW transforming growth factor-beta; TGF-beta; bone morphogenetic protein;
KW BMP; activin; anti-inflammatory; cytostatic; antiarthritic; vulnary;
KW TGF-beta modulator; wound healing; scarring; arthritis; immune response;
KW inflammatory response; tumour progression; cell proliferation; fibrosis;
KW fibrogenesis; tissue morphogenesis; tissue damage; ss.

XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 XX FT CDS 439..4410
 XX FT /*tag= a
 XX FT /product= "Human SARA-1 protein"
 XX FT /note= "Binds to receptor regulated Smad proteins"
 XX FT /transl_except= (pos:4177..4179, aa:Thr)
 XX FT /transl_except= (pos:4180..4182, aa:Gln)
 XX FT /transl_except= (pos:4198..4200, aa:Gly)
 XX FT /transl_except= (pos:4279..4281, aa:His)
 XX FT /transl_except= (pos:4306..4308, aa:Phe)
 XX FT /transl_except= (pos:4315..4317, aa:Asp)
 XX FT /transl_except= (pos:4324..4326, aa:Lys)
 XX FT /transl_except= (pos:4329, aa:Met)
 XX PN WO200005360-A1.
 XX PD 03-FEB-2000.
 XX PF 20-JUL-1999; 99WO-CA000656.
 XX PR 20-JUL-1998; 98CA-02237701.
 XX PR 10-DEC-1998; 98CA-02253647.
 XX PA (HSCR-) HSC RES & DEV LP.
 XX PI Wrana JL;
 XX DR WPI; 2000-182691/16.
 XX DR P-PSDB; AAY44749.
 XX PT New Smad Anchor for Receptor Activation (SARA) polypeptides, useful for
 XX PT developing agents for treating e.g. wounds, arthritis, fibrosis,
 XX PT inflammation, tumors, fibrogenesis or tissue damage.
 XX PS Claim 7; Page 50-51; 93pp; English.
 CC The present sequence is the cDNA encoding the human Smad Anchor for
 CC Receptor Activation protein-1 (hSARA1), a TGF-beta modulator. It is
 CC isolated from lambda ZAP human fetal brain cDNA library and has 62%
 CC identity to hSARA1. The hSARA1 proteins bind to receptor-regulated Smad
 CC proteins and ensures its appropriate localisation for activation by a
 CC Type I receptor of a TGF-beta, activin or bone morphogenetic protein
 CC (BMP) signalling pathway. SARA proteins has anti-inflammatory,
 CC cytoskeletal, anti-arthritis and vulnary activity. The SARA proteins are
 CC useful for developing agents for prevention or treatment of disorders
 CC involving TGF-beta superfamily member signalling pathway. Such agents can
 CC be used for wound healing, scarring, arthritis and fibrosis (e.g., liver
 CC and kidney), in modulating inflammatory and immune responses, tumour
 CC progression, cell proliferation or fibrogenesis and in tissue
 CC morphogenesis. It is also used for protecting, restoring and regenerating
 CC tissues after tissue damage
 XX SQ Sequence 4839 BP; 1460 A; 965 C; 1140 G; 1274 T; 0 U; 0 Other;
 Query Match 8.5%; Score 560.6; DB 3; Length 4839;
 Best Local Similarity 60.8%; Pred. No. 5.9e-100;
 Matches 990; Conservative 0; Mismatches 619; Indels 18; Gaps 4;
 3294 CCTAATGATGAGGACAGTTGCCCCACTCTCGTTGCATCTGAGAAAAGGATCATGTG 3353
 2824 CTTATTCCTGAAGATGGCCCTTCCCTCCATCTCATCTCCACTGGTGTAAAGG---AGAC 2880
 3354 CCTGTAGTAGAAGAACATCCATCTCATGAGCAGATCATTTTGTCTTGTGAAGTGAAGGC 3413
 2881 TATGCTGTGGAAGAGAAACCATCAAGATTTTCAGTAATGCGACGATTTGGAGGATGGTGGC 2940
 3414 TTTTCATCTGTATATTTGCTTAATGCTAATCTACTCGTGAATGCAATTCATATTT 3473
 2941 CCTGACCCACTGTATTTGTTTAAATGCAAAATTTGTTGTCATGTTAAATGTGTAAT 3000

QY 3474 TATTCCTCAGACAAATATTGGTACTTTTCAACCAATGATTCAGTGTGGGACAGGCA 3533
 DB 3001 TATGTGAACAGGAAGTGTGTGTTTACACCAAGGGAATGCATGCAGTGGTCACTCT 3060
 QY 3534 GAAATTATTATTCTATTGTTATGTTGGCAAAATGAAGATCTATTCTTAAGGACATCTTC 3593
 DB 3061 GAGATAGTCAATCTCTTACAGTGTTTACCGGATGAAAAGTGTTCGCCAAAGGATATCTTT 3120
 QY 3594 AGACTATTATTACACCATATATAAGGATGCTCTAAAGGAAATACATAGAAAATCTTGAC 3653
 DB 3121 AATCACTTTGTGAGCTTTATCGGAGTCTCTGCGAGGGAATGTGGTGAACATCTGGGA 3180
 QY 3654 AATATTACCTTTACTAGAGATTTTCTCAGTAGCAGGATCAGGAGATTCCTCTTTATT 3713
 DB 3181 CATTCCTTTCTCAGTCAAAAGTTTCTTGGCAGTAAAGAACATGCTGGATCTTTATATGTG 3240
 QY 3714 ACACCTACTTTTCAGAAACTTGTATGCTCTCATTAACCAAGTATCTTTTCTTTTGGGA 3773
 DB 3241 ACATCTACTACCACTCACTCAAGACCTAGTACTCCCAACCCCACTTACTCTTTTGGG 3300
 QY 3774 ATTCTTATCCAGAACCTTGAGATTCCTTGGGCAAGGTTTTTCTATCGGTTTAAATGTG 3833
 DB 3301 ATTCTTATCCAGAAATGGAACCTCTTGGGCTAAAGTATTTCTATCCGTCATGTGTG 3360
 QY 3834 AGATTGGGTGAGAAATATAAAGCATATCTCTCTCTTAACAAGCATCAGAGGCCGAAA 3893
 DB 3361 AGACTTGGAGCTGAATATCGACTTTATCCATGCCCACTTAATCACTGTGTCAGATTTCGGAAG 3420
 QY 3894 CCTCTTTTGGAGAAATAGGACACACTATTATGAATTTACTTTGTGTGACCTTCGAAATATAC 3953
 DB 3421 CCATTGTTTGGAGACCGGGCATACCATCAATGAATCTTCTTTCAGACTTCAGAAATATAC 3480
 QY 3954 CAGTATACCTTGCATAATATAGATCAACTTTGATTTCATATGAAATGGGAAAAAGCTGC 4013
 DB 3481 CAGTATACCTTGGCCAGTAGTTTCAAGTTTGGTGGTGTATATGGAAGTTCGGAAAATAGC 3540
 QY 4014 ATAAAAATACACGGAAGAAAGTACAGTATGATGAAGTACTAAATTTCTTCAATGAG 4073
 DB 3541 ATCAAAATTCACGACACAGATACATGAGATGATGAAGCCATGAACAATCCATGAG 3600
 QY 4074 CATGTCAATAGCATTTGGAGCAAGTTTTCAGTACAGAACGAGATTTCTCATCTAGTCTGTATA 4133
 DB 3601 CATGTCTTGGCAGGAGGTGCTGCTTCAATGAAAAGGCAGACTCTCATCTTGTGTGTGTA 3660
 QY 4134 CAGA---ATGATGGAATTTTGAACACAGCCACAGTGCACATGCCATCCCTAGAAA 4190
 DB 3661 CAGAATGATGATGGAACATATCAGACCAGGCTATCAGTATTTCAATCAGCCCAAGAAA 3720
 QY 4191 GTGACAGGTGCAAGTTTGTGTATTCAATGGAGCTCTAAAAACATCTTCAGGATTTCTT 4250
 DB 3721 GTGATGTTGCCAGTTTCTTGTGTTCACTGGCGCTCTGAATCTCTCTTCGGATACCTT 3780
 QY 4251 GCTAAGTCCAGCATAGTTGAAGATGGCTTAATGTTAATAAATACCTCCAGAACCATGAAT 4310
 DB 3781 GCCAAGTCCAGTATTTGTGGAAGATGTTGTTATGTTCCAGATTACTGACAGAAATGGAT 3840
 QY 4311 GCCTTTCGCGTAGCTTTACGAGAACAGAAAGACTTTTAAAAATTAATGTTGGAAAAGTTGAT 4370
 DB 3841 TCCTTTGAGCGAGGCACTCGGAGAGATGAAGACTTCACCATCACTGTGGGAAGCGGAC 3900
 QY 4371 GCAGTAGACCTGAGAGAAATACGTGGATATCTGTGGGTAGATGCTTGAAGAAAAGGAAAC 4430
 DB 3901 GCGGAGGAACCCAGGAGACATCCACATCCAGTGGGTGGATGATGACAAGAACGTTAGC 3960
 QY 4431 AAAGAGTTATCAGTTTCACTGGTGAATATCATTTACAAGGATTTTCCAAAGTGAATAATA 4490
 DB 3961 AAGGTTGCTGAAGTCCCTATAGATGGGAAGTCCATGGAGACTATTAACAATGTGAAGATA 4020
 QY 4491 AAACTGGGAAGCAGATTTTGAACCCGATGAGAAGATTTGTAATAATGTACCGAGGTGTTCTAC 4550
 DB 4021 TTCCATGGATCAGATATATAAGCAAAATGGAAGAAATGAAATCAGATGGACAGAGGTGTTTTTC 4080
 QY 4551 TTTCTAA-----AGACCAGGATTTATCTATTTTATCAACTTCTTATCAGTTTGTCA 4601

Db 4081 CTAGAAACGATGACGACCAATGGCTCAGTCACTCCTCAGATCACAGTACAGTACT 4140
 Qy 4602 AAAGAAATAGCATGGCTTGTAGTGTGCGCTGTGCGCTCACCCTGAAACCTCTAAAGT 4661
 Db 4141 GAGCATGTGCGAAAGCTTTTGGCTTGTCTGTCTCCTCACCCTGAAACCTCTCTGAAAGAA 4200
 Qy 4662 AATGGGATGATAAATTTGGACTCAGAGTTTCCATTGACACTGATATGTTGTAATTCAG 4721
 Db 4201 GATGGAATGACCAACTGGGACTACGTGTGACACTTGACTCAGATCAGGTGTGCTATCAA 4260
 Qy 4722 GCAGGATCTGAAGCCCACTTGTCCCTCAGCATTTATCTAAATGATCTTGATGCTCTG 4781
 Db 4261 GCAGGGCAATGCCAGCCCTTCCCTCGCAGTACATGAATGATCTGGATAGCGCTTG 4320
 Qy 4782 ATACTGTGATCCATGGTGGGACTCCCACTCTAGT---TTACCATTTAGAAATAGAAATTA 4838
 Db 4321 GTGCCGTGATCCATGGAGGGCCCTGCCAGCTTAGTGAGGGCCCGTGTGTCATGGAATC 4380
 Qy 4839 GTGTTTTTTCATTATAGAACATCTTTTATTAGTGAAGAAATGTGCCATATATACATATTGCAA 4898
 Db 4381 ATCTTTTATATCTGGAACATCGTATAAACAGAGAGACTTCAITTTTTTCTGTTTCAG 4440
 Qy 4899 CCTAATT 4905
 Db 4441 ACTTGTT 4447

RESULT 8

ADC64443
 ID ADC64443 standard; DNA; 1977 BP.

XX AC

XX AC

XX 18-DEC-2003 (first entry)

XX Human SARA_v5 coding sequence.

XX Hepatotropic; vaccine; protein-protein interaction;

XX Transforming Growth Factor beta; TGF beta; hepatitis;

XX Selected Interacting Domain; SID; bait; human; SARA_v5; gene; ds.

XX Homo sapiens.

XX WO2003045990-A2.

XX 05-JUN-2003.

XX 26-NOV-2002; 2002WO-EP013866.

XX 26-NOV-2001; 2001US-0333348P.

XX 31-MAY-2002; 2002US-0384537P.

XX 30-OCT-2002; 2002US-0422471P.

XX (HYBR-) HYBRIGENICS.

XX Legrain P, Gauthier J, Colland F, Jacq X;

XX WPI; 2003-505185/47.

XX P-PSDB; ADC64239.

PT New complex between two interacting proteins, useful for screening

PT molecules that inhibit transforming growth factor beta (TGF beta) or TGF

PT beta super-family of cytokines pathway for diagnosing or treating TGF

PT beta diseases or disorders.

XX Disclosure; SEQ ID NO 10; 148pp; English.

XX The present invention relates to protein-protein interactions and

XX complexes involved in Transforming Growth Factor (TGF) beta disorders

XX and/or diseases. The complex between two interacting proteins is useful

XX for screening molecules that inhibit TGF beta for diagnosing or treating

XX diseases or disorders involving TGF beta e.g., hepatitis. To illustrate

CC the invention, Selected Interacting Domains (SID) of proteins and their

CC coding sequences were isolated (ADC64243-ADC64318). Sequences ADC6434-

CC ADC64242 are bait sequences used to isolate the SIDs of the invention.

XX SQ Sequence 1977 BP; 538 A; 433 C; 485 G; 521 T; 0 U; 0 Other;

Query Match 8.4%; Score 556.2; DB 9; Length 1977;

Best Local Similarity 61.2%; Pred. No. 3.5e-99;

Matches 973; Conservative 0; Mismatches 598; Indels 18; Gaps 4;

Qy 3294 CTTATGATGAGGACAGTTTGGCCCCCACTTCTGTTGTCATCTGGAGAAAGGATCAGTG 3353

Db 391 CTTATTCCTGAAAGTGGCTTCTCCCACTTCATCTCCACTGGTGTAAAGG---AGAC 447

Qy 3354 CCGTAGTAGAAGAACATCCATCTCATGAGCAGATCAITTTGCTTCTTGAAGGTGAAGGC 3413

Db 448 TATGCTGTGGAGAGAAACCATCACAGATTCAGTAATGCAGCAGTTGGAGATGGTGC 507

Qy 3414 TTTTCATCTGTTTACATTTGTCCTAAATGCTAATCTACTCGTGAATGTCAAAATCATATT 3473

Db 508 CCGACCCCACTTGTATTTGTTTTAAATGCAATTTGTTGTCAATGTTTAAATTTGTAAT 567

Qy 3474 TATTCCTCAGACAAATATTGTTACTTTTCAACCAATGGATTGCATGGCTTGGGACAGCA 3533

Db 568 TATGTGAACAGGAAGTGTGGTGTTCACACCAAGGAATGCATGCGAGTGGTCACTCT 627

Qy 3534 GAAATTAATTAATTTGTTATGTTTGGCCAAATGAAGATACTTAATTCCTAAGGACATCTTC 3593

Db 628 GAGATAGTCATTTCTTACAGTGTTTACCGGATGAAAGTGTTCGCCAAGGATATCTTT 687

Qy 3594 AGACTATTATCACCATATATAAGGATGCTCTAAAGGAAATATACATAGAAACCTTGAC 3653

Db 688 AATCACTTTGTGCAAGCTTTTATCGGATGCTCTGGCAGGGAATGTGGTGAACACTTGGGA 747

Qy 3654 AATATTACCTTTACTGAGAGTTTCTCAGTACCAAGGATCAAGGAGATTCCTGTTTATT 3713

Db 748 CATTCCTTCTCAGTCAAGTTTCTTGGCAGTAAAGACATGATGGTATCTTATATGTG 807

Qy 3714 ACACCTACTTTTCAGAACTTGTATCTCTCATTTACCAAGTAATCCITTTCTTTGTGGA 3773

Db 808 ACATCTACCTACCACTGTCAGAACCTTAGTACTCCCAACCCCACTTACTTGTGTGG 867

Qy 3774 ATTCTTTATCCAGAAGCTTGAGATTCCCTGGGCAAGGTTTTCCTATGCTGTTTAAATGTTG 3833

Db 868 ATCTTTATCCAGAAATGGGAACTCTCTGGGCTAAAGTATTTCCCTATCCGCTGATGTTG 927

Qy 3834 AGATTGGTGCAGAAATATAAGCATATCTGCTCTCTAACCAAGCATCAGAGGCCGAAAA 3893

Db 928 AGACTTGGAGCTGAATATCGACTTTTATCCATGCCACTATTTCAGTGTCTCAGATTTCCGGAAG 987

Qy 3894 CCTCTTTTGGAGAAATAGGACACACTATTATGAATTTACTTGTGTTGACCTTCGAAATTAC 3953

Db 988 CCATTGTTTGGAGAGACGGGGCATACCATCATGAATCTTTCTTTCGAGACTTCAGAAATTAC 1047

Qy 3954 CAGTATACCTTGCATAATATAGATCAACTGTTGATTTCATATGAAATGGAAAAAGCTGC 4013

Db 1048 CAGTATACCTTGCAGTAGTTTCAAGTTTGTGGTGTGATATGAAAGTTCGAAAACTAGC 1107

Qy 4014 ATAAAAATACACAGGAAAGTACAGTGAATGAATGAAGTAAATTTCTTCCATGAG 4073

Db 1108 ATCAAAATTTCCAGCAACACAGATACATGAGATGATGAAGCCCATGAACAGTCCCAATGAG 1167

Qy 4074 CATGTCAATTTAGCATTTGGAGCAAGTTTTCAGTACAGAGCAGAGATTCTCATCTAGTCTGTATA 4133

Db 1168 CATGCTCTGGCAGGAGTGCCTGCTTCAATGAAGGAGAGACTCTCACTTCTGTGTGTA 1227

Qy 4134 CAGA---ATGATGGAATTTATGAACACAGGCCAACAGTGCACCTGGCCATCCTTAGAAAA 4190

Db 1228 CAGAAATGATGATGGAACACTATCAGACCCAGGCTATCAGTATTCACAATCAGGCCAGAAAA 1287

Qy 4191 GTGCACAGGTGCAAGTTTGTGTGTTTCAATGAGCTCTAAAAACATCTTTCAGGATTTCTT 4250

Db 1288 GTGACTGGTGCCAGTTTCTTTGTGTTTCACTGGCGCTCTGAAATCTCTTCTGATACCTT 1347

QY 3616 AGGATGCTCTAAAGGAAAATACATAGAAAACTTGGCAATATTACCTTTACTGAGAGTT 3675
DB 121 AGGATGCTCTAAAGGAAAATACATAGAAAACTTGGCAATATTACCTTTACTGAGAGTT 180
QY 3676 TTCTCAGTAGAAGATCAGGAGGATTCCTGTTTATACACCTTACTTTTCAGAACTTG 3735
DB 181 TTCTCAGTAGAAGATCAGGAGGATTCCTGTTTATACACCTTACTTTTCAGAACTTG 240
QY 3736 ATGATCTCTCATTAACCAAGTAATCTTTTCTTGTGGAATTCCTTATCCAGAGCTTCGAGA 3795
DB 241 ATGATCTCTCATTAACCAAGTAATCTTTTCTTGTGGAATTCCTTATCCAGAGCTTCGAGA 300
QY 3796 TTCCCTGGGCAAGGTTTTTCTATGCGTTTAAATGTTGAGATCGGGTCGAGAATATAAG 3855
DB 301 TTCCCTGGGCAAGGTTTTTCTATGCGTTTAAATGTTGAGATCGGGTCGAGAATATAAG 360
QY 3856 CATATCTCTGCTCTTACCAAGCATCAGAGCGCGGCAAACTCTTTTGGAGAAATAGGAC 3915
DB 361 CATATCTCTGCTCTTACCAAGCATCAGAGCGCGGCAAACTCTTTTGGAGAAATAGGAC 420
QY 3916 ACACATATTATGAACCTTACTTCTGACCTTCGAAATTTACCAGTATACCTTTGCATA 3969
DB 421 ACACATATTATGAACCTTACTTCTGACCTTCGAAATTTACCAGTATACCTTTGCATA 474

RESULT 11

AAH81649/c
ID AAH81649 standard; DNA; 455 BP.
AC AAH81649;
XX
DT 21-SEP-2001 (first entry)
DE Human differential transcription-associated cDNA SEQ ID 158.
KW Differential transcription; human; rat; tumour cell; cytostatic;
KW Ras modulator; Class II tumour suppressor gene; gene therapy; ss.
XX Homo sapiens.
OS
PN WO200157058-A2.
XX
PD 09-AUG-2001.
XX
PF 31-JAN-2001; 2001WO-EP001003.
XX
PR 31-JAN-2000; 2000DE-01004102.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;
PI Grips M, Hellriegel M, Schmitz A, Sers C;
XX
DR WPI; 2001-483415/52.

XX Nucleic acids differentially expressed between tumor and normal cells,
PT useful for diagnosis or therapy of tumors and for screening active
PT agents.

XX Disclosure; Page 372; 579pp; German.

XX This invention describes a nucleic acid (I) with differential expression
CC between tumour and normal cells and which has cytostatic activity. (I)
CC work as modulators of Ras activity by inducing expression of tumour
CC suppressor genes (I), and polypeptides encoded by them, are useful as
CC targets for diagnosis or therapy and in screening to determine the
CC effects of an active compound (potential pharmaceutical) on a cell line,
CC particularly for diagnosis and treatment of tumors, especially by
CC modulating expression of (I) (by gene therapy, antisense RNA or ribozyme
CC methods) or by modulating the amount and/or location of (I)-encoded
CC polypeptides (by administration of the polypeptide or its activator,
CC antibody (optionally as a conjugate) or inhibitor). The method allows
CC identification of many Class II tumour suppressor genes (i.e. genes that

CC are not primary targets for tumour-initiating mutations). AAH81492-
CC AAH82376 represent the human and rat derived nucleic acid fragments
CC described in the method of the invention
XX
SQ Sequence 455 BP; 185 A; 61 C; 70 G; 139 T; 0 U; 0 Other;
Query Match 6.7%; Score 444; DB 5; Length 455;
Best Local Similarity 99.8%; Pred. No. 2.4e-77;
Matches 455; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 5337 GACTTATATATCTTTTGAATAAAGCTGATACTTATTTGTAATTTGGAGTGGAGACCTAC 5396
DB 455 GACTTATATATCTTTTGAATAAAGCTGATACTTATTTGTAATTTGGAGTGGAGACCTAC 396
QY 5397 CTCCATAATTTAGATAAACTCTTTTGGATTATAATCAGAAATTTTCCCTTTCTTCTCA 5456
DB 395 CTCCATAATTTAGATAAACTCTTTTGGATTATAATCAGAAATTTTCCCTTTCTTCTCA 336
QY 5457 AATTATACATATGATGATTTATATATATCCACATATATAGTTTCCCTGATTAATGGAT 5516
DB 335 AATTATACATATGATGATTTATATATATCCACATATATAGTTTCCCTGATTAATGGAT 276
QY 5517 ATTAATAAATTTGCGGGTCTTCAGGACTTTTGGCTTCTATATTTAAGTATATTTT 5576
DB 275 ATTAATAAATTTGCGGGTCTTCAGGACTTTTGGCTTCTATATTTAAGTATATTTT 216
QY 5577 ATAGCAAGAACATATCTTGAATGTTTATAAATCTTTTAAATCTTTTATATAGTGAATAT 5636
DB 215 ATAGCAAGAACATATCTTGAATGTTTATAAATCTTTTAAATCTTTTATATAGTGAATAT 156
QY 5637 TTTTGTATCACAAATGCAATTTTTCCTCTCTCTTCCCTTCCAACTATACCACTGAT 5696
DB 155 TTTTGTATCACAAATGCAATTTTTCCTCTCTCTTCCCTTCCAACTATACCACTGAT 97
QY 5697 TACCACCTTCTAAGAGTGACTGACGACGGGCCAGATGACCTTTGAAGTAGTCATTATGTAG 5756
DB 96 TACCACCTTCTAAGAGTGACTGACGACGGGCCAGATGACCTTTGAAGTAGTCATTATGTAG 37
QY 5757 CAATAAATGAAGCCTGAAACAGGTTTTTTTACTTCC 5792
DB 36 CAATAAATGAAGCCTGAAACAGGTTTTTTTACTTCC 1

RESULT 12

ACH20104
ID ACH20104 standard; cDNA; 391 BP.
XX
AC ACH20104;
XX
DT 13-OCT-2003 (first entry)
DE Human adult lung cDNA #1107.
XX
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW Genome mapping; biodiversity; genetic disorder.
XX Homo sapiens.
OS
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;

Db 301 TAAATGACTCTAAATCG 317

RESULT 14
ABX48202
ID ABX48202 standard; cDNA; 393 BP.

XX
AC ABX48202;
DT 21-FEB-2003 (first entry)
XX

DE Bovine EST associated with lactation/muscle/fat deposition #13367.
XX
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX
OS Bos Taurus.
XX

PN US2002137139-A1.
XX
PD 26-SEP-2002.
XX

PF 24-SEP-2001; 2001US-00960352.
XX
PR 12-JAN-1999; 99US-0115707P.
PR 11-JAN-2000; 2000US-00480902.
XX

PA (BYAT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX

PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
DR WPI; 2003-110599/10.
XX

PT New nucleic acid associated with lactation, and muscle and fat
PT deposition, useful for genome mapping, gene identification and analysis,
PT cattle breeding, or for genetically improving cattle.
XX

PS Claim 2; SEQ ID NO 13367; 245pp; English.
XX

CC The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMFD), derived from
CC cattle, and the LMFD nucleic acid can specifically hybridize to a second
CC nucleic acid molecule comprising any of 15112 nucleotide sequences,
CC appearing as ABX4836-ABX49947, or complements of them. Also included are
CC ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
CC acid linked to a promoter and a 3' non-translated sequence that
CC functions in the cell to cause termination of transcription and addition
CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC (2) determining a level or pattern of a molecule in a bovine cell or
CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC of the 15112 nucleic acid sequences or its complement or fragment) with a
CC complementary nucleic acid molecule obtained from the bovine cell or
CC tissue, where hybridisation between the marker nucleic acid and the
CC complementary nucleic acid permits the detection of the molecule; and (b)
CC detecting the level or pattern of the complementary nucleic acid, where
CC the detection of the complementary nucleic acid is predictive of the
CC level or pattern of the molecule. The LMFD nucleic acid is used for
CC determining a level or pattern of a molecule in a bovine cell or tissue.
CC It is useful for genome mapping, gene identification and analysis, cattle
CC breeding, preparation of constructs for use in cattle gene expression, or
CC for genetically improving cattle. The present sequence is one of the
CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
CC present sequence was not shown in the specification but was obtained in
CC electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?docID=20020137139
XX

SQ Sequence 393 BP; 124 A; 53 C; 63 G; 153 T; 0 U; 0 Other;
XX

Query Match 3.0%; Score 199.6; DB 7; Length 393;

Best Local Similarity 81.4%; Pred. No. 1.9e-29;
Matches 311; Conservative 0; Mismatches 49; Indels 22; Gaps 6;

QY 5111 TTCTTTTAAAGAAATTTATAGCATTTTACTGTGTTATTTTAAATGCT-AAGCCAAAGTATCT 5169
DB 22 TGTCTTATAGAAATTTTGTAGCATTTTCTGTGTTTAAATGCTTAAAGCCAAAGTATCT 81
QY 5170 GCATTTAGGTATACCTCTTTATGCCAATAAATGATTTTAAATGAAGCTCTTTTCAGATGTA 5229
DB 82 GCACT-----TTATGCCAAGAACGATTTTATAGTGAAGGCTCTTTTCAGATGTA 129
QY 5230 ACCTTATGAGGAATATCTGCTTTTGTGTATATGCCAGTTAGCACTAGTAATCTTCTTAAAGT 5289
DB 130 ACCTTATGAGGAATATCTGCTTTTGTGTATATGCCAGTTAGCACTAGTAATCTTCTTAAAGT 189
QY 5290 CTG-TCAAATTTGATTTTCAGTGGCACAACAAACAGTTTGTAGGCTTTAGACTT-----AT 5343
DB 190 CTGTTAAATGATTTTCAATGGCAC--AGACCAGTTTTCAGATCTTACACTTATATGAT 247
QY 5344 AATCTTTGAATAAACTGA-TAACTTATTTGTATATTAATGAGTGGAGACCTACCTCCAT 5402
DB 248 AATCTTTGAATAAAGCAGATTAGCTTATTTTATATGAGGTTAAATTTGAGTGGAGACCTACCTCCAT 307
QY 5403 AATTAGATAAAGCTCTTTTGGATTTATTAATCAGAAATTTTGCCTTTTCTTCAAAATTAT 5462
DB 308 GATTAGATAAAGATTTTGGCATTAAATCAGAAATTTTGCCTTTTCTTCAAAATTAT 367
QY 5463 TACATATGATGATTTATATAT 5484
DB 368 AATTTAATATATATATATATAT 389

RESULT 15
ABL28217
ID ABL28217 standard; DNA; 4068 BP.
XX
AC ABL28217;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 36124.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 36124; 21pp + Sequence Listing; English.
XX

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4068 BP; 1067 A; 1021 C; 1110 G; 870 T; 0 U; 0 Other;
Query Match 2.9%; Score 191.8; DB 4; Length 4068;
Best Local Similarity 50.0%; Pred. No. 1.1e-27;
Matches 620; Conservative 0; Mismatches 577; Indels 42; Gaps 4;
QY 3653 CAATATTACCTTTACTGAGAGTTTCTCAGTAGCAAGGATCAGGAGGATTCCTGTTTAT 3712
DB 2817 CACCAGTCCGGAATCCCAATTTCTGGGTCCCGGAGCATGGAGGATTTATCTTTAT 2876
QY 3713 TACACCTACTTTTCAGAACTTGATGATCTCTCATACCAAGTAATCCCTTTCTTTGTGG 3772
DB 2877 TCGGCCAACGTACAGGTGCTTGAGGGAGTATTTGACAGACAAATCCGTATCTGTTGG 2936
QY 3773 AATCTTTATCCAGAGCTTGAGATTCCTCGGCAAGGTTTTTCCATGCGTTTAATGTT 3832
DB 2937 CGTCTCATCCATAGACACGAGGTACCTTGGGCTAAGGTGCTACCTTTGGGTTGATCCT 2996
QY 3833 GAGATTGGGTGCAGAAATTAAGCATATCTGCTCTTAACAGCATCAGAGGCGGAA 3892
DB 2997 CCGCTTGGGTGCCAGATCGGTACTATCCCTGTCCATTGATCTCGGTGCGCGAGGGA 3056
QY 3893 ACCCTTTTGGAGAAATAGGACACACTATTATGAACTTACTTTGTTGACCTTCGAAATTA 3952
DB 3057 GTCGGTGATGGCGAATAGTCAAACTATCATTTAACTTTCTGTGGATTTTCGCAACTA 3116
QY 3953 CCAGTATACCTTGTCATTAATATAGATCAACTGTTGATTCATATGGAATGGGAAAGCTG 4012
DB 3117 TTCGTACTCTCTACCGGCCATCCGAGGATTAATATCCATGGAGGATAGGCGAGCGAC 3176
QY 4013 CATAAATAATACCGGAAAGAGTACAGTGTATGTAATGAAAGTACTAAATTTCTTCCAATGA 4072
DB 3177 GGTGATTAATCCAAAGTACCGCCAGCAGGATGTAATTAAGGCATCAATATGCTAGTGA 3236
QY 4073 GCATGTCATTAGCAATGGAGCAAGTTTCAGTACAGAGAGAGATTCATCTAGTCTGTAT 4132
DB 3237 TCATATACCTGCTTTTGTGCAACTTCTCAAAAGTGGCCGATGGTCACTTGGTTTGAT 3296
QY 4133 ACAGANTGATGGAATTTATGAA-----ACACAGGCCCAACAGTGC 4171
DB 3297 GCAGATATATCAACGACGATCGATCCGAGATGTAATCTCTCCACGCGGCGGATTAATAT 3356
QY 4172 CACTGGCCCATCTAGAAAGTGACAGGTGCAAGTTTGTGGTATTCATGGAGCTCTAAA 4231
DB 3357 TCAGGGTCAACCCAGGAAATCACTGGAGCTAGCTTCTTTGTGCTAAATGAAGCCCTCAA 3416
QY 4232 AACATCTTCAGGATTTCTTGCTAAGTCCAGCATAGTTGAAGATGCTTAATGGTACAAAT 4291
DB 3417 GAGCAGCAGTGGCTTGTGGGTAAGTCAGCATCTGGAGGATGGCCCTAATGGTACAGAT 3476
QY 4292 AACTCCAGAGACCATGAATGGCTTGGGCTAGCTTTAGAGAACAGAAAGACTTTTAAAT 4351
DB 3477 TATGCTGCCAAATGGAGGAGTGGCTGAGCGCTTGGCAACCAAAACGACATCGACAT 3536
QY 4352 TACATGTCGGAAGTTGATG---CAGTAGACCTGAGAGATACGTTGATATCTCTGGGT 4408
DB 3537 CGTCTGTGGAACCATTTGACGCCACCGACGATCAGAGTGAGATTTGAAGCATTAAGTGGT 3596
QY 4409 AGATGCTCAAGAAAGAAACAAAGAGGTTATCAGTTTCAGTGGATGAATATCATTTACA 4468
DB 3597 GGACAAATATCGTGACATCATGTTGGGGGTTAAGTCGCCCAATGATGACGAGCCCATGGA 3656
QY 4469 AGGATTTCCAAAGTGAAAAA---TAAAACTGGAAGCAGATTTTGAACCCGATGAGAAGAT 4525
DB 3657 TGGCAATTTCAACACATGCGAGTGTGAAATGGCCAGCTTTAACTATTCCAATACCAACTA 3716
QY 4526 TGTAATAATGTACCGAGGTGTTCTACTTTCTTAAAGGACGAGGATTTATCTATTTATCAAC 4585

DB 3717 TGCCATCCGCTGAGTATATCTACATTTAAGTGCAGGACTGGTATGCAACAATGG 3776
QY 4586 TTCTTTAT-----CAGTTTGCAAAAAGAAATAGCCATGGCTTGTAGTGTGC 4630
DB 3777 TGGAAACTGTGAGGAGATAACACGGATTTCGGAGCAACTGCTCGCAGTGCCTCGATGGC 3836
QY 4631 GCTGTGCCCTCACCTGAAAACTCTAAAAAGTAATGGGATGAATAAATTTGACTCAGAGT 4690
DB 3837 TCTGTGCCACTTCTTCGATCTTTTAGCAGCTGCTGGCAITTAATAGCTTTGGCCTAAGAGC 3896
QY 4691 TTCCATTGACACTGATATGTTGAATTTTCAGGCAGGATCTGAAGGCCAACTTTCTGCTCA 4750
DB 3897 CACGCTCGATCAGATACGCTGGGCTATGAAGCCGAGCTCGTGGCTCCAAGCTACCGCC 3956
QY 4751 GCATTAATCTAAATGATCTTGATAGTGTCTCTGATACCTGTGATCCATGTTGGACCTCCAA 4810
DB 3957 GTTGATCATGAACGGCTGGCAACCACTCTGATAGCCACGCTGCATGCGGAATCCTCGG 4016
QY 4811 CTCTAGTTTACCATTAGAAATAGAAATAGTGTGTTTTTCAT 4849
DB 4017 CATTCAGAGCCCATCATCTTAGAGCTGATCTTTTACAT 4055

Search completed: May 17, 2004, 18:15:14
Job time : 1534.9 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 16:26:01 ; Search time 10163.4 seconds
(without alignments)
19486.266 Million cell updates/sec

Title: US-09-744-167-3
Perfect score: 6632
Sequence: 1 actccgcggcggttagctc.....ttggaggcacatttgaagt 6632

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estmu.*
- 5: em_estov.*
- 6: em_estpi.*
- 7: em_estro.*
- 8: em_hcc.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_hcc.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: em_gss_hum.*
- 18: em_gss_inv.*
- 19: em_gss_pln.*
- 20: em_gss_vrt.*
- 21: em_gss_fun.*
- 22: em_gss_mam.*
- 23: em_gss_mus.*
- 24: em_gss_pro.*
- 25: em_gss_rod.*
- 26: em_gss_phg.*
- 27: em_gss_vrl.*
- 28: gb_gss1.*
- 29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1695	25.6	3164	11	AK080779 Mus muscu
2	951.8	14.4	1035	12	BM925782 AGENCOURT
3	918.8	13.9	1973	11	AK045090 Mus muscu
4	801.8	12.1	1120	13	BQ430548 AGENCOURT

RESULT 1
AK080779
LOCUS
DEFINITION
Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length enriched library, clone:B130024H06 product:hypothetical protein, full insert sequence.
ACCESSION
AK080779.1 GI:26099473
VERSION
AK080779.1
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
MEDLINE
PUBMED
10349636
REFERENCE
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
PUBMED
11042159

ALIGNMENTS

AK080779 3164 bp mRNA linear HTC 20-SEP-2003
Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length enriched library, clone:B130024H06 product:hypothetical protein, full insert sequence.
AK080779.1 GI:26099473
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
MEDLINE
PUBMED
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
PUBMED
11042159

5	759.8	11.5	882	12	BI767442
6	747.6	11.3	836	13	BQ431811
7	741.8	11.2	897	13	BQ84268
8	739	11.1	740	9	BM700151
9	712	10.7	723	12	BM976393
10	708.8	10.7	727	9	AU133002
11	687.2	10.4	717	9	AL596824
12	682.6	10.3	935	14	CD557100
13	679.4	10.2	713	14	CB306662
14	676.4	10.2	711	10	AW953665
15	668.4	10.0	758	14	CB962678
16	665.4	10.0	800	14	CB309758
17	645.4	9.7	658	13	EX641326
18	644	9.7	645	13	EX485406
19	643	9.7	758	14	CD101985
20	638.4	9.6	641	13	BU078916
21	626.8	9.5	822	13	BQ224046
22	623.2	9.4	949	12	BG287850
23	621.2	9.4	862	13	BQ428538
24	609.6	9.2	914	10	BE883109
25	609	9.2	609	14	CB215048
26	602.4	9.1	620	14	CB297891
27	602.2	9.1	747	13	BQ571837
28	597.2	9.0	658	14	CB450372
29	595.4	9.0	601	9	AL703681
30	593.4	8.9	606	12	BM971938
31	583.4	8.8	586	14	CA949552
32	583	8.8	889	10	BF038696
33	579.6	8.7	609	14	CD676502
34	569.4	8.6	582	14	CA949827
35	567	8.5	568	9	AV717110
36	566.8	8.5	583	9	AI989276
37	564	8.5	575	12	BM052837
38	563.8	8.5	579	10	BE222531
39	563	8.5	567	10	BF439324
40	561	8.5	674	12	BG546318
41	560.2	8.4	579	12	BI964158
42	555.8	8.4	608	14	CD239102
43	555	8.4	782	12	BG287961
44	553.4	8.3	567	9	AI435398
45	552.8	8.3	677	14	CF950835

3
REFERENCE
AUTHORS
Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuina,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Tashiro,H., Itoh,M.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,K., Izawa,M., Ohara,E., Watanabe,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Murakami,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384 format
sequencing pipeline with 384 multiplexed sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED
11076861
4
REFERENCE
AUTHORS
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
REFERENCE
AUTHORS
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3164)
6
REFERENCE
AUTHORS
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imokani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurahara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saioh,R., Saichoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tgami,M., Tagawa,A., Takahashi,F., Takaku-Akai,H.,
Takeda,Y., Tanaka,T., Tomaru,A., Toyota,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. Tomohiro Kono (Department of Animal
Science, Tokyo University of Agriculture, 1737 Hunaiko Atsugi City,
Kanagawa Prefecture, Japan) whose assistance we gratefully
acknowledge.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/
7
FEATURES
source
1. 3164
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:BI30024H06"
/db_xref="MG:2410429"
/db_xref="taxon:10090"
/clone="BI30024H06"
/tissue_type="parthenogenote"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="9.5 days embryo"
1. 3164
misc_feature
/note="hypothetical protein (evidence: ProCrest)"
ORIGIN

```
Db 1367 -----GGGGTCCCTTGGTTTATATACACTACTCTCCAGAAACCCCT 1406
Qy 3734 TGTATCTCTCATTACCAAGTAATCCTTTTCTTTGTGGAAATCTTATCCAGAAGCTTGA 3793
Db 1407 TGTATGCTCTCCCGTCCAGGAGTCTCTTCTCTCGCGCATCTTATCCAGAAGCTGA 1466
Qy 3794 GATTCCTCGGG-CAAAAGGTTTTCTTATGCTTAAATGTTGAGATGGGTGCGAGATATA 3852
Db 1467 GATTCCTCGGCCAAAGGTTTTCTTATGCGGTAAATGTTGCGATTTGGGAGAGAAATAC 1526
Qy 3853 AAGCATATCTGCTCTCTCTAA-CAAGCATCAGAGGC-CGAAAAACCTCTTTTGGAGAAAT 3910
Db 1527 AAGCATACCTGCTCTCTCTAA-CAAGCATCAGAGGC-CGAAAAACCTCTTTTGGAGAAAT 1586
Qy 3911 AGGACACACTATTATGAACTTACTTGTGACCTTGGAAATACAGATACCTTGTGATTA 3970
Db 1587 AGGACATACCTATTATGAACTTACTTGTGATCTTGGAAATACAGATACCTTGTGATTA 1646
Qy 3971 TATAGATCAACTGTTGATTCATATGGAATGGGAAAGCTGCATAAAAATACACGGAA 4030
Db 1647 CATCGATCAGCTATTGATTCATATGGAATGGGAAAGCTGTATAAAAATACCTCGAA 1706
Qy 4031 AAAGTACAGTGTATGTAAGAACTACTTAAATCTTCCAAATGAGATGTCATTAGCATTTGG 4090
Db 1707 AAAGTACAGTGTATGTAAGAACTACTTAAATCTTCCAAATGAGATGTCATTAGCATTTGG 1766
Qy 4091 AGCAAGTTTCAGTACAGAGAGATCTCATCTAGTCTGTATACAGATGATGAAATTA 4150
Db 1767 AGCTAGTTTTAGTACAGAGAGATCTCATCTGTGCTGTGTACAGAGTATGGAGTTTA 1826
Qy 4151 TGAACACAGGCCAACAGTGCCTAGCCCTCTAGAAAAAGTGACAGGTGCAAGTTTGT 4210
Db 1827 TCAGACTCAGGCCAACAGTGCCTAGCCCAAGGAAAGTGACAGGTGCAAGTTTGT 1886
Qy 4211 GGTATTCAGTGGAGCTCTAAAAACATCTTCAGGATTTCTTGTGTAAGTCCAGATAGTTGA 4270
Db 1887 GGTGTTTAAATGGTCTCTGAAACATCTTCAGGATTTCTTGGCAAAATCCAGCATAGTGA 1946
Qy 4271 AGATGGCTTAAATGTAACAATAACTCCAGAGACCATGAATGGCTTGGCGCTAGCTTTACG 4330
Db 1947 AGATGGCTTAAATGTAACAATAACTCCAGAGACCATGGAGGCTTGGCGCTAGCTTTACG 2006
Qy 4331 AGAACAGAAAGACTTTTAAATATCATGTGGAAAGTTGATGCGATAGACCTCGAGAGATA 4390
Db 2007 AGACAAAGAGACTTTAGAAATCCAGTGTGGAAAGTTGATGCGGTAGACCTCGAGAGATA 2066
Qy 4391 CGTGGATATCTGCTGCTAGATGCTGAGAAAGAGAAACAAAGGAGTTATCAGTTCCAGT 4450
Db 2067 TGTGGACATCTGTTGGGTGGATTGAGAGAGAGAGAAACAAAGGTTATCAGTTCCAGT 2126
Qy 4451 GGATGGAAATATCATTAAGGATTTCCAAGTGAAGAAATAAACCTGGAAGCAGATTTTGA 4510
Db 2127 AGATGGAAATGTCAGTGGAGGATTTCCAAGTGAAGAAATAAACCTGGAAGCAGATTTTGA 2186
Qy 4511 AACCGATGAGAAATGTAAGATGACCGAGGTTCTTACTTTCTAAAGGACCGAGTTT 4570
Db 2187 AACTGAGAGAGAGACTGTGAAATGCTAGGTTCTTACTTCTTAAAGGACCAAGATAT 2246
Qy 4571 ATCTATTTTATCAACTTCTTATCAGTTTGGCAAGAAATAGCCATGGCTGTAGTGTCTGC 4630
Db 2247 ATCTATTTTATCAAGTAGTACAGTTTGGCAAGAGATGCTGTGCTGTAGTGTCTGC 2306
Qy 4631 GCTGTGCCCTCAGCTGAAACCTTAAAGTAATGGAATGGAATGGAATGGAATGGAATGGAAT 4690
Db 2307 CCTGTGCCCTCAGCTGAAACCTTAAAGTAATGGAATGGAATGGAATGGAATGGAATGGAAT 2366
Qy 4691 TTCCATTGACACTGATGTTGATTTCCAGGAGGATCTGAAGGCCAATCTTCTGCTCA 4750
Db 2367 GTCCATTGACACTGATGTTGATTTCCAGGAGGCTGTGAAGGCCAATCTTCTGCTCA 2426
Qy 4751 GCATTATCTAAATGATCTTGTATGCTCTGATACCTGTGATCCATCGTGGGA-CCTCCA 4809
```

```
Db 2427 GCATTACCTGAATGATCTGGATAGTGCCCTGATCCCTGTGATCCACGGCGGACCCCTCA 2486
Qy 4810 ACTCTAGTTTACCATTAGAAATAGAAATAGTGTGTTTTCATTATAGAAATCTTTTCTTGT 4869
Db 2487 ACTCTAGTCTGCCACTAGAAATAGAAATAGTGTGTTTTCATTATAGAAATCTTTCTGAGT 2546
Qy 4870 GAAAGAAATGTCATATATACATATTTGCAACCTAAATTTGTTTAAACCTAACTCCAGCACTAA 4929
Db 2547 GAAACCGTGTACTGTGGTGTATTTTGCAAACTGATGGGTAAACCGAATCCAGCACTAA 2606
Qy 4930 AGCTGAATGTCACAAACACCTAAAGATATAAATATGCTGATTTTGAACACATAGCT 4989
Db 2607 AACGGAATGTCACAAACACCTAA-----ATGTTTGATAATTGAAATATATAATTT 2656
Qy 4990 TTGCTCTTTAGCGAGATGATCTTTTCAAAATCATAGCAATAATTTAAATATCTAAAA 5049
Db 2657 TCGTTTTTATGAGTGAAGATGATTTTAAAAATCATAGCAAGATTTTAAAGTCTCTAAAA 2716
Qy 5050 ATTTAAGATCCATACCTTTCTGTAGCTTCAAAATTAATTAAGTACTAAAAAGCAAGG 5109
Db 2717 ATTTAAG-GAATCATCAITTTTATAGCTTTATGCTTTATTAAGAACTAAAGAGG-AAAT 2773
Qy 5110 ATTTCTTTTAAAGAAATTTATAGCAATTTTACCTGTTTAAATGCTAAAGCAAGATCT 5169
Db 2774 ATTTCTAACTTAAATTTTACAGTATTTTACTGTGTACCTTAACTGTTAAAGCCAAAGCACT 2833
Qy 5170 GCACCTTAGTATACCTCTTTTATGCCAATAATGATTTTAAATGAAGGCTCTTTTCAGATGTA 5229
Db 2834 GGCCTTAGCCCTACTCTCTCATAGATGACGA----- 2863
Qy 5230 ACCTTATGAGAAATATCTGCTTTGTGTATATGCGAGTAGAACTAGTGTCTTAAAGT 5289
Db 2864 ---GAATGAAGAGAGAGACCTGCGTTGTGTGGATGCTAGTGAGAACACTGGTTT-TGAAGT 2919
Qy 5290 CTGTCAAAATGTATTTTCACTGGGCACAAAAACCAAGTTTGGAGTCTTAGACTTTATAATCT 5349
Db 2920 CTGTTTAAATGTATCTCAGTAAGAGACACAGTTTCAAGATCTTAGACTTTATAATCT 2979
Qy 5350 TTGAATAAACTGATAACTTATTGTATTAATTTGGAGTGGAGACCTACCTCCATATATGA 5409
Db 2980 TTGAATAAACTGATAACTTATTGTATAAATGGAATGAAGACCTACCTCCATGATTACA 3039
Qy 5410 TAACTCTTTTGGATTATAATCAGAAATTTTCCCTTTTCTTCTCAAAATTTACATAT 5469
Db 3040 TACAACTGTTT-GATTAAATTTAGAAATTTTCCCTTTTATTTCTTCTCAAAATTTCAATAT 3098
Qy 5470 GTATGTTATATATATCCACATATATAGTTTCCCTGATTTAAATGGAATTTAAAAATATG 5529
Db 3099 CTATATA-TATATATACACATATAAATATATATATTTTCTG---ATTAAAAATATG 3154
Qy 5530 CGGTGCTT 5538
Db 3155 CGGTGCTT 3163
```

RESULT 2
BM925782
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BM925782 1035 bp mRNA linear EST 12-VAR-2002
AGENCOURT_6710339 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5764025
5', mRNA sequence.

BM925782
BM925782.1 GI:19376161
EST.

Homo sapiens (human)
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1035)

NIH-MGC <http://img.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM12816 row: m column: 18
 High quality sequence stop: 794.
 Location/Qualifiers
 1..1035
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5764025"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_114"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
 male brains, age range 23-27 yo. Library is oligo-dT
 primed and directionally cloned (EcoRV site is destroyed
 upon cloning). Average insert size 1.5 kb, insert size
 range 1-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 019. Note:
 this is a NIH_MGC Library."

FEATURES

source

ORIGIN

Query Match 14.4%; Score 951.8; DB 12; Length 1035;
 Best Local Similarity 96.8%; Pred No. 7.8e-179;
 Matches 1003; Conservative 0; Mismatches 28; Indels 5; Gaps 3;

QY 5277 TGGTTCTTAAAGTCGTCAAAATGTATTCAGTGGCAAAAACACAGTTTGGAGTCTTA 5336
 Db 1 TGGTTCTTAAAGTCGTCAAAATGTATTCAGTGGCAAAAACACAGTTTGGAGTCTTA 60

QY 5337 GACTTATATCTTTTGAATAAACTGATACTTATTTGTATAATTTGAGTGGAGACCTAC 5396
 Db 61 GACTTATATCTTTTGAATAAACTGATACTTATTTGTATAATTTGAGTGGAGACCTAC 120

QY 5397 CTCATAATTAGATAAACTCTTTTGGATTATATCAAGATTTTGCCTTTTCTTCTCA 5456
 Db 121 CTCATAATTAGATAAACTCTTTTGGATTATATCAAGATTTTGCCTTTTCTTCTCA 180

QY 5457 AATTATTACATATGATGTATATATATATCCACATATAGTTTCCCTGATTAAATCGAT 5516
 Db 181 AATTATTACATATGATGTATATATATATCCACATATAGTTTCCCTGATTAAATCGAT 240

QY 5517 ATTTAAATAATTGGGGTCTTCAGGACTTTTTCCTCTATATTTAAGTATATTTT 5576
 Db 241 ATTTAAATAATTGGGGTCTTCAGGACTTTTTCCTCTATATTTAAGTATATTTT 300

QY 5577 ATACAGAAATATTCGAATGTTTATATAATCTTTAATAATTTATATAGTATATAT 5636
 Db 301 ATACAGAAATATTCGAATGTTTATATAATCTTTAATAATTTATATAGTATATAT 360

QY 5637 TTTTGTATCAAGTCATTATTTTTCCTTCCTTCCTCCAACTATACCACTGTATT 5696
 Db 361 TTTTGTATCAAGTCATTATTTTTCCTTCCTTCCTCCAACTATACCACTGTATT 419

QY 5697 TACCATTCTTAAGTAGTACTGACGACGGGCCAGATGACCTTGAAGTAGTCATTATAG 5756
 Db 420 TACCATTCTTAAGTAGTACTGACGACGGGCCAGATGACCTTGAAGTAGTCATTATAG 479

QY 5757 CAATAAATGAGCTGGAACAGGTTTCTTACTTCCACTTTAATCCCTTAGAAATTTCTTG 5816
 Db 480 CAATAAATGAGCTGGAACAGGTTTCTTACTTCCACTTTAATCCCTTAGAAATTTCTTG 539

QY 5817 GCAACTTCGCATATTTTCAATGACACGGTGTATAGTATAAATTTAAATGAACATTA 5876
 Db 540 GCAACTTCGCATATTTTCAATGACACGGTGTATAGTATAAATTTAAATGAACATTA 599

QY 5877 CTTTGGCATATTTTAAATCTTATATATGGTAGTATTTTATATAACAGGATATTACATA 5936

Db 600 CTTTTCATATTTTAAATCTTATATATAGTATTTTATATAACAGGATATTAAACATA 659

QY 5937 AGTTAAATCCTATGTTTGAATTTGTTACAGAGCTTTCTCTTTTACTTCAACAGCAAA 5996

Db 660 AGTTAAATCCTATGTTTGAATTTGTTACAGAGCTTTCTCTTTTACTTCAACAGCAAA 719

QY 5997 AAAGTGGGGGCATATTTAGTCTCTGTCATTTTAAGTTATGTAAAAAATTTATCATTT 6056

Db 720 AAAGTGGGGGCATATTTAGTCTCTGTCATTTTAAGTTATGTAAAAAATTTATCATTT 779

QY 6057 TTGATGCTTTAAACATTTCTCATGTTGTAATATATATTTTGTATCAAAACACTCATAT 6116

Db 780 TTGATGCTTTAAACATTTCTCATGTTGTAATATATATTTTGTATCAAAACACTCATAT 839

QY 6117 TTCAAGAAAAAGAAATTTATGTTTAAATAGCCCTGTTTAAAGAAAAATTTTATGAAGCATC 6176

Db 840 TTTCAGAAAAAGAAATTTATGTTTAAATAGCCCTGTTTAAAGAAAAATTTTATGAAGCATC 899

QY 6177 TCAACTTGAAGTCAAGTCAAAAGTTATTAACCTCAGGATCTGAGGTCTCAAGCTAGGAGAGA 6236

Db 900 TCCACTTGAAGTCAAGTCAAAAGTTATTAACCTCAGGATCTGAGGTCTCAAGCTAGGAGAGA 959

QY 6237 CTGAGAAATTTTAAATCAGTTTGGGCATA---TAGTTTGGACTGAATCAGAT-CTGTAGTAC 6292

Db 960 CTGAGAAATTTTAAATCAGTTTGGGCATAATTTAGTTTGGACTGAATCAGATCTCTGGAGAAC 1019

QY 6293 TTAGCCAAAGACAATT 6308

Db 1020 TTAGCCAAAGACAATT 1035

RESULT 3
 AK045090
 LOCUS
 DEFINITION
 Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
 enriched library, clone:B13031L15 product:weakly similar to
 ENDOPIN [Homo sapiens], full insert sequence.
 AK045090
 VERSION AK045090.1 GI:26337062
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 PUBMED 10349636
 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 PUBMED 11042159
 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
 Yoneda, Y., Iehikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 PUBMED 11076861
 4
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.	
	Functional annotation of a full-length mouse cDNA collection	
	Nature 409, 685-690 (2001)	
	5	
	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.	
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
	Nature 420, 563-573 (2002)	
	6 (bases 1 to 1973)	
	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hangaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashida, K., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	
	Direct Submission	
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp).	
	URL: http://genome.gsc.riken.go.jp/ , Tel: 81-45-503-9222, Fax: 81-45-503-9216	
	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.	
	Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	
	Tissues were provided by Dr. Tomohiro Kono (Department of Animal Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City, Kanagawa Prefecture, Japan) whose assistance we gratefully acknowledge.	
FEATURES	Please visit our web site for further details.	
	URL: http://genome.gsc.riken.go.jp/	
	URL: http://fantom.gsc.riken.go.jp/ .	
	Location/Qualifiers	
	1. 1973	
source	/organism="Mus musculus"	
	/mol_type="mRNA"	
	/strain="C57BL/6J"	
	/db_xref="FANTOM:DB:B130031L15"	
	/db_xref="MGI:2410449"	
CDS	/db_xref="taxon:10090"	
	/clone="B130031L15"	
	/tissue_type="parthenogenote"	
	/clone_lib="RIKEN full-length enriched mouse cDNA library"	
	/dev_stage="9.5 days embryo"	
ORIGIN	210. .>1973	
	/note="unnamed protein product; putative weakly similar to ENDOFIN [Homo sapiens] (SPTR AAL30772, evidence: PASTY, 58%ID, 38.9%length, match=1764)"	
	/codon_start=1	
	/protein_id="BAC32215.1"	
	/db_xref="GI:26337063"	
Query Match	13.9%; Score 918.8; DB 11; Length 1973;	
	Best Local Similarity 73.5%; Pred. No. 2.8e-172;	
	Matches 1351; Conservative 0; Mismatches 412; Indels 75; Gaps 11;	
	208 AGCATACAGAATTAATCTGATPAAGTCTGCAGGTAGGACAGATTATTTAAAG 267	
	169 AGACATAGAGAATTCAGTTCTGAGTAAGCCACAGGAGGATGGACAGTTACTTTAAAG 228	
Db	268 CAGCTCTCAGTGACTTGGACAACTCCTTGATGATTTTGAACAGACACCAGATGAACAAG 327	
	229 CAGCTCTCAGTGCTTGGACAACTTCTTGATGATTTTGAACAGATCCAGATGAACAAG 288	
	328 ATTATCTCCAGATGTPACAAAATGTCATATGATTCTAACCACTGCTCAGTTTCTTCAGAGT 387	
	289 ATTATCTCCAG-----ATGCATATGCTTTTAAACCAGTCTCAGTTTCTTCAGAGT 339	
	388 TGGCTTCTCCACAGCAACTTCATCTCCCAAGACCAAGAGTGCCTTAATAGTTTGTG 447	
Qy	340 CGGCCCTCTCCACAGCTAGCT---TTACTGTCAAAGGACCAAGATGCATCAGTACTTGTG 396	
	448 CCTCATCAGAAAACAAGCTATGGA---ACAAATGACAGTTTCCCTGAATGAAAAACACTCA 504	
	397 CCTCATCAGAGCATGCTCTGGAAGATGCAATGAGACTTTTCTTGAAGGAAGATACACG 456	
	505 AGGACTTACTTCTATACAAAATGAAAAATGTAAACAGACTTGATCTTCTTCTTCTG 564	
	457 AGGGCTTAACCTCTAGACCGAATGAAAAAAGCTAGCGGAGCTTGATCTTCTTCTTCTG 516	
Db	565 TGGATCGTGTACTTCCAGATGAAATCCAGCCGTATATATATGACAGATGTAGTAAACCTTA 624	
	517 TGGATCGGAGTACTTCCAGATGAGATCCAGCCCTTCTGTATGAGACGGTGTAGTAGCCTG 576	
	625 TCTGTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 684	
	577 TCTGTGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 636	
	585 TTAATAAATATGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 744	
Qy	637 TTAAGCAGTATGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 696	
	745 TATCTTTCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 804	
	697 CGTCTTCAGTTCAGAGGCTCTCAGTCTCTTTCAGTGTGATGATGATGATGATGATGATG 756	
	805 GAGAACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 864	
	757 GAGAGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 813	
Db	865 AATTGGGTATAAAGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 924	
	814 AATTGGGTATAAAGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 873	
	925 ATTAAAGATATAAAGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 984	
	874 ATTAAAGATATAAAGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 930	
	985 CTGCTTTGACTCGACAAAGTTCACAAATGTTTCATGCCCAAGACAGCTTACCAACAAGA 1044	
Qy	931 CTGCTTTGACTCGACAAAGTTCACAAATGTTTCATGCCCAAGACAGCTTACCAACAAGA 990	
	1045 GCCAGCATGTGGATTACTTAAAGATGTTGGCTGTAGTAAAGAGGAGTAGATGTTGGCAG 1104	
	991 GGCTGCCATGTGAAATTTGTTTAAAGATGATGCTGTTTGGCAGAGAGAGTAGCTGTGG 1050	
	1105 TCATACTCCCGCAGATGTTTAAAGAGAGG-----GCAAGACAGTGTCTTGCCT 1158	
	1051 CAGTCAATAATACAGATGTTTAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1110	
Db	1159 GAGCCCTTCGCAAAATGAAGATTTATGCTTTAAATGATTTCAAATCAAGAGATGAAAAAT 1218	
	1111 GCAAGCTTCCAGAAATGAAGGATATCCCAAGTGCACCTTCAAAAGATGAAAAAT 1170	

JOURNAL
COMMENT

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.
Email: cgaapb-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L14M13211 row: i column: 07

High quality sequence stop: 656.

Location/Qualifiers

FEATURES

source

1. .1120

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6015438"

/tissue_type="embryonal carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH MGC 92"

/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally; oligo-dT primed.

Average insert size 2.5 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC Library."

ORIGIN

Query Match 12.1%; Score 801.8; DB 13; Length 1120;

Best Local Similarity 91.0%; Pred. No. 5.4e-149;

Matches 954; Conservative 0; Mismatches 78; Indels 16; Gaps 9;

QY 1219 TCAATATTACCTGACTTTTCCCTTTCAGGAGATAAGACT-----GTTATTAACAACATCTG 1272

DB 1171 TCAAGTTACCTGACTTTTCCCTTTCAGGAGATAAGACTTTGTTTCATGAACAGACTG 1230

QY 1273 CACAGAAGACTCAAAAAGTTTACACCTTAAGGATATGATGTAATCAAGATTCCTCTT 1332

DB 1231 TAAAGAAGACTCAAGAAATTTAGACCTTAAGATATTAATGATAT----- 1276

QY 1333 CAGCTTTACATGTTTCCAGTAAAGATGTCGGTCTCATTTCTCTGCTCTCTGCTGCTG 1392

DB 1277 ---AGTCCATGTTTCAGGTGATGATGTGCCCTTCACTGCTGCTGCTTTTCAATGTCTG 1332

QY 1393 GGTCTATGCTGGATCATTAATTAAGAGTAAGACAGCGGGTGATTTTACCTCAGCATG 1452

DB 1333 GGTCTTTGTGATGCTGTTAATTAACAATTAAGACAGATGATTTTACCTCTAATG 1392

QY 1453 AACATAAAGATAATATACAAGATGACGTACTATACATGAAGAATAACAGAACAGTTG 1512

DB 1393 AGTCTGAAGGCCAGAACAAATGATGACGTACTATACATGAAGATACAGAAAGATGATG 1452

QY 1513 TTCTAGTGGGGAACCATCAAGAGATGATCTTTTGAACAGGAAATGTAAGACCA 1572

DB 1453 TTTTAGATGGG-----GAGACAGATCTTTCAAGAGAGAAACGTGTAGGAGCA 1500

QY 1573 TACTCTTCAGTCATTAATTTGAAGGATGGAAGACAGAAAGATAGATCCTGACCAAGACAG 1632

DB 1501 TATTTCTTCAGCCAGTAATGAAGAAGAGGGGAGGAAAGGTTGAGGTTGAGGATGG 1560

QY 1633 TAATCAGAGCTGAGTCTTTTGA-----TGTTGGTGACACAGATTTCTACAGTTTGAAT 1686

DB 1561 TAATCAGTGGTGAGTCTTTGAGATGCCCTGAGGAGCCGACGCTGCTGCGCAGCTGGT 1620

QY 1687 CTCAGAGGGGCTTTCTGCACTCATGTCCTGAGAGTCTTCTGATGTTGTGAGAGTTTA 1746

DB 1621 CTCGGTGCACTTTCTGCTGCAAGTGTCCAGAGGCTCTCTGTTCCATGTGAAGGTCT-- 1678

QY 1747 TTAATTAATTTTCAAGCAATGATATGATGGGCAAGACTTAGATTAATTAATTTGATG 1806

DB 1679 ---CACCTTTTCCAGCAGTGTATGATGGGCAAGAGTATAGATTACTTTAATATTGATG 1734

QY 1807 AAGGCGAAAGTGGCCCACTAATAGTGTGCTGACTGATGCTGCTGCTGCTGCTGACAGAC 1866

DB 1735 AAAGCATGAGAGTGGCATACTAATAGCATGCTGAACTCGATGCTTTCTGAGGAAAC 1794

QY 1867 AGTATCTTCAGACCACTAACATAAAGTCTTTTGAAGAAATGTAATGACTCTAAATGCG 1926

DB 1795 AGTGCTTTTGAACCTTAACACATGCTGCTGGGAAATGTAATGATTCCTCCAGTTGC 1854

QY 1927 AAATGAATCAGATAGATGAAGCTTAGATGATGGAACATCAATAATATATATTTCA 1986

DB 1855 AGATGAATCAGATCACTATGAAGGACTACATGATGAATAATGCTGGGATATATACTTCA 1914

QY 1987 ATGCAAGAGCAGGAGCTATTGGGAAAGTCATGTTATT 2024

DB 1915 ATGCTGAAGCAGGAGCTGCTCGAGAAATGGAGGTGTT 1952

RESULT 4

BQ430548

LOCUS

DEFINITION BQ430548 1120 bp mRNA linear EST 24-MAY-2002

5', mRNA sequence.

ACCESSION BQ430548

VERSION BQ430548.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1120)

NIH-MGC <http://mgi.nci.nih.gov/>

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

```

|||||
657 CCATGATAGTACTGTACAGAACACAGATGATACAGTCTCTCAATTAACAAATAG 716
|||||
845 AGAAATCGAGGAGTCAAGAAATGGGTATAAAAGTAGATACACAC-TTTCAGATTCT 903
|||||
717 AGAAATCGAGGAGTCAAGAAATGGGTATAAAGTAGATACACACTTTTCAGATTCT 776
|||||
904 ATAAATACAGTGAACACAG-AAAATTTAAAGATAAAAGATCTTTAATCAGTTAGATCA 962
|||||
777 ATAAATACAGTGAACACAGAAAAATTTAAAGATAAAAGATCTTTAATCAGTTAGATCA 836
|||||
963 ATTGTTGATTTAAACATGTCATCTGTTTACCTCGACAAAGTCCAAAATGTTTCATG-C 1021
|||||
837 TTTGTTGATTTTACATGTCATCTGTTTGAATCCCAAAGTCCAAAATGTTTCATGCC 896
|||||
1022 CAAAGACAAGCTACACACAG-AGCCAGCCATGT-GGATTAATAAAGATGTGG- --- 1075
|||||
897 CAAAGACCAGCTACACCCAGAACCCAGCAGCATGTGGATTACTAAAGATGTTGGGCTT 956
|||||
1076 -CTTAGTAAAGAGGAGTATGATGTGGCAGTCACTAAGTCCG-CAGAAATGTTTAAAGAA 1133
|||||
957 AATTAAAGAGGAGTATGATGTGGCATTCTTACCTCGCCCAAATGTTTAAAGAAA 1016
|||||
1134 GAGGG---CAAGACAAGTCTTTGACCTGCAGCCTTCCGAAAAATCAAG-ATTATGCT 1188
|||||
1017 GAAGGCAAGACAGGTGCTTTTAACTCGACCTTCCGAAAGAGAAATATATGCC 1076
|||||
1189 TAAATGATCAAAATCAAGAGATGAAAA 1216
|||||
1077 TTAAGGGTCCAAATCCAGAAAAAGAAA 1104
|||||

```

RESULT 5

```

BI767442
LOCUS      892 bp      mRNA      linear      EST 25-SEP-2001
DEFINITION 603057595F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5207230 5',
mRNA sequence.

```

```

ACCESSION  BI767442
VERSION     BI767442.1
KEYWORDS    GI:15759020
SOURCE      Homo sapiens (human)

```

```

ORGANISM    Homo sapiens

```

```

REFERENCE   1. (bases 1 to 892)
AUTHORS     NIH-MGC http://mgi.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            plate: LLAM1519 row: m column: 23
            High quality sequence stop: 788.

```

FEATURES

```

source

```

```

1..892
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5207230"
/lab_host="DH10B"
/clone_lib="NIH_MGC_122"
/notes="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site 1: NotI; Site 2: EcoRV (destroyed); RNA source:
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size

```

RESULT 6

```

BQ431811

```

range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC Library."

ORIGIN

```

Query Match      11.5%; Score 759.8; DB 12; Length 882;
Best Local Similarity 95.6%; Pred. No. 1.2e-140;
Matches 825; Conservative 0; Mismatches 32; Indels 6; Gaps 4;

QY 4716 TTTCCAGGAGGATCTGAAGGCCAACTCTGCTCAGCATTAATCTAAATGATCTTGATAGT 4775
DB |||||
QY 4776 GCTCTGATACCTGTGATCCATGCTGGGACCTCCAACTCTAGTTTACCATTAAGAAATAGAA 4835
DB |||||
QY 61 GCTCTGATACCTGTGATCCATGCTGGGACCTCCAACTCTAGTTTACCATTAAGAAATAGAA 120
QY 4836 TTAGTGTTTTCATTATAGAACATCTTTTGTAGTGAAGAGATGCGCATATACATATG 4895
DB |||||
QY 121 TTAGTGTTTTCATTATAGAACATCTTTTGTAGTGAAGAGATGCGCATATACATATG 180
QY 4896 CAACCTAAATTTGTTAAACTAACTCCAGCACATAAGCTGAAATGCCACAAACACATAAAG 4955
DB |||||
QY 181 CAACCTAAATTTGTTAAACTAACTCCAGCACATAAGCTGAAATGCCACAAACACATAAAG 240
QY 4956 TATAAATATGCTGATTTTGAACACATAGCTTGTCTTTTAGGCGAGGATGATCTTT 5015
DB |||||
QY 241 TATAAATATGCTGATTTTGAACACATAGC-TTGCTCTTTTAGGCGAGGATGATCTTT 299
QY 5016 TCAATCATATTAGCACAAATATTAAATATCTAAATTTAAAGAGATCCATCTCTCTGTAG 5075
DB |||||
QY 300 TCAATCATATTAGCACAAATATTAAATATCTAAATTTAAAGAGATCCATCTCTCTGTAG 359
QY 5076 CTTTCAATTAATTTAAGTACTAAAGACAGAGATTTCTTTTAAAGAAATTTATAGCAAT 5135
DB |||||
QY 360 CTTTCAATTAATTTAAGTACTAAAGACAGAGATTTCTTTTAAAGAAATTTATAGCAAT 419
QY 5136 TACTGTGTTATTAAATGCTTAAGCAAGATCTGCACCTTAGGTATACCTCTTTATGCCA 5195
DB |||||
QY 420 TACTGTGTTATTAAATGCTTAAGCAAGATCTGCACCTTAGGTATACCTCTTTATGCCA 479
QY 5196 ATAATGATTTTAATGAAGGCTCTTTTCAGATGTAACTTATGAAGGAAATATCTGCTTTG 5255
DB |||||
QY 480 ATAATGATTTTAATGAAGGCTCTTTTCAGATGTAACTTATGAAGGAAATATCTGCTTTG 539
QY 5256 TGTATATGCCAGTTAGATAGTCTTTTAAAGCTGTCAAAATGTATTTTCAGTGGCACA 5315
DB |||||
QY 540 TGTATATGCCAGTTAGATAGTCTTTTAAAGCTGTCAAAATGTATTTTCAGTGGCACA 599
QY 5316 AAAACCACTTTTGAGGCTCTTAGACTTATAATTTCTTGAATAAACTGATAAATTTTGT 5375
DB |||||
QY 600 AAAACCACTTTTGAGGCTCTTAGACTTATAATTTCTTGAATAAACTGATAAATTTTGT 659
QY 5376 ATAATGAGTGGAGACCTACTCCATTA-TTAGATAAATCTTTTGGATTAATCAAG 5434
DB |||||
QY 660 ATAATGAGTGGAGACCTACTCCATTAATTAATTAATTAATTAATTAATTAATCAAG 719
QY 5435 AATTTTGCTTTTCTTCTCAAAATTTTACATATGATGATATATATATCCACATATAT 5494
DB |||||
QY 720 AATTTTGCTTTTCTTCTCAAAATTTTACATATGATGATATATATATCCACATATAT 778
QY 5495 AGTTTTCCTGATTAATAGGATATTAATAATTAATTCGGGGTCTCTCAGACATTTTGTTC 5554
DB |||||
QY 779 AGTTTTCCTGATTAATTAATTAATTAATTAATTCGGGGTCTCTCAGACATTTTGTTC 838
QY 5555 TATATTTAAGTATATGTTTATA 5577
DB |||||
QY 839 T---TTTAAAGTATATTTGTATA 858

```

LOCUS BQ431811 836 bp mRNA linear EST 24-MAY-2002
 DEFINITION AGENCOURT 7907398 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6154816
 5', mRNA sequence.
 ACCESSION BQ431811
 VERSION BQ431811.1 GI:21170887
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 836)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM13496 row: h column: 17
 High quality sequence stop: 508.
 Location/Qualifiers
 1..836
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6154816"
 /tissue_type="retinoblastoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_67"
 /note="Source: eye; Vector: pCMV-SPORT6; Site1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.75 kb. Library constructed by Life
 Technologies."
 ORIGIN
 Query Match 11.3%; Score 747.6; DB 13; Length 836;
 Best Local Similarity 96.7%; Pred. No. 3.3e-138;
 Matches 784; Conservative 0; Mismatches 25; Indels 2; Gaps 2;
 QY 3127 CTACTTCTGGTTCATTACACTAGATGATGATGTTTTCAGAACTGAAGAACCATCTA 3186
 DB 1 CTACTTCTGGTTCATTACACTAGATGATGATGTTTTCAGAACTGAAGAACCATCTA 60
 QY 3187 GTCTACTGGTGTCTAGTTAAACAGCAATTTACCTATTGCTAGTATTTCAGATTATAGT 3246
 DB 61 GTCTACTGGTGTCTAGTTAAACAGCAATTTACCTATTGCTAGTATTTCAGATTATAGT 120
 QY 3247 TACTGTGTGATTAACCAAGTATGTCGCAATAGATTAGTCTTCTACCTAAATGATCAGG 3306
 DB 121 TACTGTGTGATTAACCAAGTATGTCGCAATAGATTAGTCTTCTACCTAAATGATCAGG 180
 QY 3307 ACAGTTTCCCCCACTTCTGGTTCATCTGGAGAAAGGGATCAGTCCCTGTAGTAGAAG 3366
 DB 181 ACAGTTTCCCCCACTTCTGGTTCATCTGGAGAAAGGGATCAGTCCCTGTAGTAGAAG 240
 QY 3367 AACATCCATCTCATGAGACATCTTTTCTTCTTGAAGTGAAGGCTTTCATCTCTGTTA 3426
 DB 241 AACATCCATCTCATGAGACATCTTTTCTTCTTGAAGTGAAGGCTTTCATCTCTGTTA 300
 QY 3427 CATTTGTCCTAAATGCTAATCTACTCGTGAATGTCAAATTCATATTTTATTCCTCAGACA 3486
 DB 301 CATTTGTCCTAAATGCTAATCTACTCGTGAATGTCAAATTCATATTTTATTCCTCAGACA 360
 QY 3487 AATATTGCTATTTTCACCAATGATGATGCTGGACGCGGAGAAATATTATTC 3546
 DB 361 AATATTGCTATTTTCACCAATGATGATGCTGGACGCGGAGAAATATTATTC 420
 QY 3547 TATTGTTATGTTTGCCAAATGAAGATACTATTTCCTAGGACATCTTCAGACTATTATCA 3606

DB 421 TATTGTTAATGTTGCCAAATGAAGATACTATTCTTAGGACATCTTCAGACTATTATCA 480
 QY 3607 CCATATATAAGGATGCTCTAAAGGAAATACATAGAAAACCTTGACAAATATTACCTTTA 3666
 DB 481 CCATATATAAGGATGCTCTANAGGAAATACATAGAAAACCTTGACAAATATTACCTTTA 540
 QY 3667 CTGACAGTTTCTCAGTAGCAAGGATCAGGAGGATTCCTGTTTATTACACTACTTTTC 3726
 DB 541 CTGACAGTTTCTCAGTAGCAAGGATCAGGAGGATTCCTGTTTATTACACTACTTTTC 600
 QY 3727 AGAAAATTCAGTATCTCTCATTTACCAAGTAATCTTTTCTTTGTGGAATCTTATCCAGA 3786
 DB 601 AGAAAATTCAGTATCTCTCATTTACCAAGTAATCTTTTCTTTGTGGAATCTTATCCAGA 660
 QY 3787 AGCTTGAGATTCCTGGGCAAGGTTTCTCTATGCGTTTAAATGTTGAGATGGGTGCAG 3846
 DB 661 AGCTTGAGATTCCTGGGCAAGGTTTCTCTATGCGTTTAAATGTTGAGATGGGTGCAG 720
 QY 3847 AATATAAGGATAT-CCTGCTCTCTTAAAGCATCAGAGG-CGAAAACCTCTTTTGTG 3904
 DB 721 AATATAAGGATATCCCTGCTCTCTTAAAGCATCAGAGG-CGAAAACCTCTTTTGTG 780
 QY 3905 AGAAAATAGGACACATATTATGAATCTTACTT 3935
 DB 781 GAGAAAATAGGACCCACTATTATGAACCTT 811
 RESULT 7
 BQ884268 897 bp mRNA linear EST 16-AUG-2002
 LOCUS AGENCOURT 8682001 Lupski_sciatic_nerve Homo sapiens cDNA clone
 DEFINITION IMAGE:6197558 5', mRNA sequence.
 ACCESSION BQ884268
 VERSION BQ884268.1 GI:22276276
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 897)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. James R. Lupski
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM13607 row: m column: 15
 High quality sequence stop: 607.
 Location/Qualifiers
 1..897
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6197558"
 /sex="male"
 /tissue_type="sciatic nerve"
 /dev_stage="adult, 70 yr"
 /lab_host="DH10B"
 /clone_lib="Lupski_sciatic_nerve"
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site1:
 NotI; Site 2: SalI; cDNA made by oligo-dT priming.
 Directionally cloned using the following adaptors:
 5'-TCGACCCAGCGCTCG-3' and
 5'-GACTAGTCTAGATCGGCGCCCT(15)-3'. Size selected >
 1 kb for average insert length 1.87 kb. This is a primary
 library, non-amplified. Library constructed by Life

FEATURES
 source

Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine) and is available through Life Technologies."

ORIGIN

Query Match 11.2%; Score 741.8; DB 13; Length 897;
Best Local Similarity 94.2%; Pred. No. 4.7e-137;
Matches 847; Conservative 0; Mismatches 42; Indels 10; Gaps 7;

QY 4489 TAAAACTGGAAGCAGATTTCGAACCGATGAGAAGATTGTAAGATGACCGAGTGTTCT 4548
DB 1 TAAAACTGGAAGCAGATTTCGAACCGATGAGAAGATTGTAAGATGACCGAGTGTTCT 60

QY 4549 ACTTCTTAAGGACACAGATTATCTATTTATCAACTTCTATCAGTTTGCAAGAAA 4608
DB 61 ACTTCTTAAGGACACAGATTATCTATTTATCAACTTCTATCAGTTTGCAAGAAA 120

QY 4609 TAGCATGGCTGTAGTCTCGCTGCTGCCCTCACCTCGAAACCTCTAAAGATGATGGA 4668
DB 121 TAGCATGGCTGTAGTCTCGCTGCTGCCCTCACCTCGAAACCTCTAAAGATGATGGA 180

QY 4669 TGAATAAAATGGACTCAGAGTTTCCATGTGACACTGATATGTTGAATTTTCAGCAGGAT 4728
DB 181 TGAATAAAATGGACTCAGAGTTTCCATGTGACACTGATATGTTGAATTTTCAGCAGGAT 240

QY 4729 CTGAAGCCCACTCTGCTCCTCAGCATTATCTTAATGATCTTGATAGTCTGCTCTGATACCTG 4788
DB 241 CTGAAGCCCACTCTGCTCCTCAGCATTATCTTAATGATCTTGATAGTCTGCTCTGATACCTG 300

QY 4789 TGATCCATGTGGGACCTCCAACTCTAGTTTACCATTAGAAATAGAAATAGTGTGTTTCA 4848
DB 301 TGATCCATGTGGGACCTCCAACTCTAGTTTACCATTAGAAATAGAAATAGTGTGTTTCA 360

QY 4849 TTATAGAACATCTTTTATAGTGAAGAATGTCATATGATATGATATGATATGATATGAT 4908
DB 361 TTATAGAACATCTTTTATAGTGAAGAATGTCATATGATATGATATGATATGATATGAT 420

QY 4909 TAAAACTAACTCCAGCACTAAAGCTGAAATGAAATGAAATGAAATGAAATGAAATGATGCT 4968
DB 421 TAAAACTAACTCCAGCACTAAAGCTGAAATGAAATGAAATGAAATGAAATGAAATGATGCT 480

QY 4969 GATTTTGAACACATAGCTTGTCTTTAGGAGGAAATGATCTTTTCAAAATCATTTAGC 5028
DB 481 GATTTTGAACACATAGCTTGTCTTTAGGAGGAAATGATCTTTTCAAAATCATTTAGC 540

QY 5029 ACAATATTAAATATCTAAAGATTTAAGAGATCCATCTTCTGATAGCTTTTCAAAAT 5088
DB 541 ACAATATTAAATATCTAAAGATTTAAGAGATCCATCTTCTGATAGCTTTTCAAAAT 600

QY 5089 TTAAGTACTAAAAGACAAGGATTTCTTTTAAAGAAATTTATAGCAATTTACTGTGTTATTT 5148
DB 601 TTAAGTACTAAAAGACAAGGATTTCTTTTAAAGAAATTTATAGCAATTTACTGTGTTATTT 660

QY 5149 AATGCTAAGCCAAGTATCTGCACTTAGTATACCTTCTTTATAGCCAAATGATTTTAA 5208
DB 661 AATGCTAAGCCAAG-ATCTGCCCTTAGTATACCTTCTTTATAGCCAAATGATTTTAA 719

QY 5209 TGAAGGCTCTTTTTCAGATGTAACCTTATGAAGAAATATCTGTTTG- TGTATATGCCA 5266
DB 720 AGAAGCTCTTTTCA- AAGAACCTTATGAAGAAATATCTGTTTG- TGTATATGCCA 778

QY 5267 GTTAGAATAC- TGTGTTCTTAAAGTCTGTC- AAATGTAATTTTCAGTGTG- CACAAAAACCA 5322
DB 779 GCTAGAATAGGGGTTTCTTAAAGTCTGCGGCAAAATTTGATTTCCGTTGGGCAAAAAACCA 838

QY 5323 GTTTTGAAGCTTAGAGCTTATAA- TTCTTTTGAATAAACCTGATACTTATTTGTATAA 5379
DB 839 TTTTGAAGCTTAAACTAAAAATTTCTTTTAAATAAACTGGAACCTTATTTGTATAA 897

RESULT 8
AL700151
LOCUS

DEFINITION DKFZp686I20117_r1 686 (synonym: hlccc3) Homo sapiens cDNA clone
DKFZp686I20117_5', mRNA sequence.

ACCESSION AL700151
VERSION AL700151.1 GI:19620684
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 740)
AUTHORS Ottenwaelder, B., Obermaier, B., Mewes, W., Mewes, H.W., Weil, B. and Wiemann, S.
TITLE EST (Ottenwaelder, B., Obermaier, B., Mewes, H.W., Weil, B. and Wiemann, S.)
JOURNAL Unpublished (2001)
COMMENT Contact: MIPS

FEATURES
source
1..740
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686I20117"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlccc3)"
/note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN

Query Match 11.1%; Score 739; DB 9; Length 740;
Best Local Similarity 100.0%; Pred. No. 1.7e-136;
Matches 739; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3742 TCTCATTAACCAAGTATCTTTTCTTTCTGGAATCTTATCCAGAGCTTGAGATCCCT 3801
DB 2 TCTCATTAACCAAGTATCTTTTCTTTCTGGAATCTTATCCAGAGCTTGAGATCCCT 61

QY 3802 GGGCAAAAGTCTTCTATGCTTTTAAATGTTGAGATGGGTGCAGAAATATAAGCATATC 3861
DB 62 GGGCAAAAGTCTTCTATGCTTTTAAATGTTGAGATGGGTGCAGAAATATAAGCATATC 121

QY 3862 CTGCTCTCTTAACAGCATCAGAGCCGAAACCTCTTTTGGAGAAATAGGACACATA 3921
DB 122 CTGCTCTCTTAACAGCATCAGAGCCGAAACCTCTTTTGGAGAAATAGGACACATA 181

QY 3922 TTATGAACCTACTCTTGACCTTCGAAATACCAATACCACTTGCATATATAGATCAAC 3981
DB 182 TTATGAACCTACTCTTGACCTTCGAAATACCAATACCACTTGCATATATAGATCAAC 241

QY 3982 TGTGATTCATATGAAATGGGAAAGCTGCATAAAAATACCAAGAAAAAGTACAGTG 4041
DB 242 TGTGATTCATATGAAATGGGAAAGCTGCATAAAAATACCAAGAAAAAGTACAGTG 301

QY 4042 ATGTAATGAAGTACTAAATCTTCCATGAGCATGCTATTAGCATTCGAGCAAGTTTCA 4101
DB 302 ATGTAATGAAGTACTAAATCTTCCATGAGCATGCTATTAGCATTCGAGCAAGTTTCA 361

QY 4102 GTACAGAGCAGATTTCTCATCTAGTCTGTATACAGAAATGATGGAATTTATGAAACACAGG 4161
DB 362 GTACAGAGCAGATTTCTCATCTAGTCTGTATACAGAAATGATGGAATTTATGAAACACAGG 421

QY 4162 CCAACAGTGCCTACTGGCCATCTCTAGAAAAAGTGACAGGTGCAAGTTTGTGTTTCAATG 4221

```

Db 422 CCAACAGTGCACCTGGCCATCCTTAGAAGAGTACAGTGCAGTTTGTGTTATCAATG 481
QY 4222 GAGCTCTAAACAACTCTTTCAGGATTTCTTCTAAGTCCAGCATAGTTGAAGATGGCTTAA 4281
Db 482 GAGCTCTAAACAACTCTTTCAGGATTTCTTCTAAGTCCAGCATAGTTGAAGATGGCTTAA 541
QY 4282 TGGTACAAATTAACCTCCAGAGACCATGATGGCTTCGGCTAGCTTTACGAGACACAGAAAG 4341
Db 542 TGGTACAAATTAACCTCCAGAGACCATGATGGCTTCGGCTAGCTTTACGAGACACAGAAAG 601
QY 4342 ACTTTAAATTAACATGTTGGGAAAGTTCATGACAGTAGACCTCGAGAGATACGTGGATATCT 4401
Db 602 ACTTTAAATTAACATGTTGGGAAAGTTCATGACAGTAGACCTCGAGAGATACGTGGATATCT 661
QY 4402 GCTGGGTAGATGCTCAAGAAAGAAAGAAACAAAGAGTTATCAGTTTCAGTGGATGGATAT 4461
Db 662 GCTGGGTAGATGCTCAAGAAAGAAAGAAACAAAGAGTTATCAGTTTCAGTGGATGGATAT 721
QY 4462 CATTACAGGATTTCCAG 4480
Db 722 CATTACAGGATTTCCAG 740

```

RESULT 9

```

BM976393 723 bp MENA linear EST 21-FEB-2003
LOCUS UI-CF-EN1-acz-h-01-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
DEFINITION UI-CF-EN1-acz-h-01-0-UI 3', mRNA sequence.
ACCESSION BM976393
VERSION 1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 723)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cDNA
sequence: 1-23, >AT rich#Low complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.

```

FEATURES

```

source
1..723
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EN1-acz-h-01-0-UI"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EN1"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a

```

```

modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-EN1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Cystic Fibrosis
Epithelial Cells. The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an
oligo-dr primer containing a Not I site. Double stranded
cDNA was ligated to an EcoR I adaptor, digested with Not
I, and cloned directionally into pT7T3-Pac vector. The
clonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CTGCTCAGGT.
TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG LIB=UI-CF-EN1
TAG_SEQ=CTGCTCAGGT"

```

ORIGIN

```

Query Match 10.7%; Score 712; DB 12; Length 723;
Best Local Similarity 100.0%; Pred. No. 4.1e-131;
Matches 712; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4815 AGTTTACCATTAGAAATAGATTAGTGTCTTTTCATTATAGAACATCTTTTGTAGTGAAG 4874
Db 723 AGTTTACCATTAGAAATAGATTAGTGTCTTTTCATTATAGAACATCTTTTGTAGTGAAG 664
QY 4875 AATGTGCCATATTACATATTGCAACCTAATTTGTTTAAACTTAACCTCCAGCACTAAAGCTG 4934
Db 663 AATGTGCCATATTACATATTGCAACCTAATTTGTTTAAACTTAACCTCCAGCACTAAAGCTG 604
QY 4935 AAATGCCACAAACACATAAAGTATAAATATGCTCTGATTTTGAACACATAGCTTTGCT 4994
Db 603 AAATGCCACAAACACATAAAGTATAAATATGCTCTGATTTTGAACACATAGCTTTGCT 544
QY 4995 CTTTGTAGGAGGAATGATCTTTTCAAAATCATTTAGCACAATATTAAATATCTTAAATAATTA 5054
Db 543 CTTTGTAGGAGGAATGATCTTTTCAAAATCATTTAGCACAATATTAAATATCTTAAATAATTA 484
QY 5055 AGAGATCCCATCTTTCTGTAGCTTTTACAAATTAATTTAAGTACTTAAAGACAGAGATTTTC 5114
Db 483 AGAGATCCCATCTTTCTGTAGCTTTTACAAATTAATTTAAGTACTTAAAGACAGAGATTTTC 424
QY 5115 TTTTAAAGAAATTTATAGCAATTTTACGTGTATTATTTAAATGCTTAAGCCAAAGATCTGCACT 5174
Db 423 TTTTAAAGAAATTTATAGCAATTTTACGTGTATTATTTAAATGCTTAAGCCAAAGATCTGCACT 364
QY 5175 TAGGTATACCTCTTTTATGCCAATAATGATTTTAAATGAAGGCTCTTTTTCAGATGTAACCTT 5234
Db 363 TAGGTATACCTCTTTTATGCCAATAATGATTTTAAATGAAGGCTCTTTTTCAGATGTAACCTT 304
QY 5235 ATCAAGAAATATCTGCTTTGTATATGCCAGTTAGATACTGTTTCTTAAAGTCTGTC 5294
Db 303 ATCAAGAAATATCTGCTTTGTATATGCCAGTTAGATACTGTTTCTTAAAGTCTGTC 244
QY 5295 AAATTTGATTTTCAGTGGCACAAAACACAGTTTGTAGGCTTTAGACTTAAATCTTTGAA 5354
Db 243 AAATTTGATTTTCAGTGGCACAAAACACAGTTTGTAGGCTTTAGACTTAAATCTTTGAA 184
QY 5355 TAAACTGTATTAATTTGTATTAATTTGGAGTGGAGACCTCTTCCATTAATTAGATAAAC 5414
Db 183 TAAACTGTATTAATTTGTATTAATTTGGAGTGGAGACCTCTTCCATTAATTAGATAAAC 124
QY 5415 TCTTTTGGATTATAATCAGAAATTTTGCCCTTTTCTTCTCAAAATTTATACATATGATG 5474
Db 123 TCTTTTGGATTATAATCAGAAATTTTGCCCTTTTCTTCTCAAAATTTATACATATGATG 64
QY 5475 TATTATATATCCACATATATAGTTTTTCCCTGATTAATTAATCGATTAATAATAA 5526
Db 63 TATTATATATCCACATATATAGTTTTTCCCTGATTAATTAATCGATTAATAATAA 12

```

RESULT 10

```

AUI33002
LOCUS       AUI33002 NT2RP4 Homo sapiens cDNA clone NT2RP4001068 5', mRNA          EST 01-AUG-2002
DEFINITION   AUI33002 NT2RP4 Homo sapiens cDNA clone NT2RP4001068 5', mRNA
ACCESSION   AUI33002
VERSION     AUI33002.1 GI:10993541
SOURCE      EST.
ORGANISM    Homo sapiens (human)
REFERENCE   1 (bases 1 to 727)
  AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  TITLE     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  JOURNAL   Helix Research Institute
  COMMENT   1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..727
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cclone="NT2RP4001068"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /clone_lib="NT2RP4"
     note            /note="vector: pME18SFL3; mRNA from NT2 neuronal precursor
                     cells after 2-weeks retinoic acid (RA) induction"
ORIGIN
Query Match      10.7%; Score 708.8; DB 9; Length 727;
Best Local Similarity 99.0%; Pred. No. 1.8e-130;
Matches 721; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
QY 4965  GTCTGATTTTGAACACATAAGCTTTGCTCTTTAGCGAGGAATGATCTTTTCAAAATCAT 5024
DB 1  GTCTGATTTTGAACACATAAGCTTTGCTCTTTAGCGAGGAATGATCTTTTCAAAATCAT 60
QY 5025  TAGCACATATTTAAATATCTMAAAATTTAAGAGATCCATCTTCTGTAGCTTTACAAT 5084
DB 61  TAGCACATATTTAAATATCTMAAAATTTAAGAGATCCATCTTCTGTAGCTTTACAAT 120
QY 5085  TAATTTAAGTACTAAAAGACAAGGATTTCTTTTAAGAAATTTATAGCAATTTACTGTGTT 5144
DB 121  TAATTTAAGTACTAAAAGACAAGGATTTCTTTTAAGAAATTTATAGCAATTTACTGTGTT 180
QY 5145  ATTTAAATGCTAGCCAAAGTATCTGCATAGTATACCTCTTTATGCCAATAATGATT 5204
DB 181  ATTTAAATGCTAGCCAAAGTATCTGCATAGTATACCTCTTTATGCCAATAATGATT 240
QY 5205  TTAATGAAGGCTCTTTTCAGATGTAACCTTATGAAGGAATATCTGCTTTGTGTATATGC 5264
DB 241  TTAATGAAGGCTCTTTTCAGATGTAACCTTATGAAGGAATATCTGCTTTGTGTATATGC 300
QY 5265  CAGTTAGAACTACTGGTTTCTAAAGCTGTGCAAAATGTTATTCATGTCGCAAAAACAGT 5324
DB 301  CAGTTAGAACTACTGGTTTCTAAAGCTGTGCAAAATGTTATTCATGTCGCAAAAACAGT 360
QY 5325  TTTCAGGCTCTAGACTTATATCTTTTGAATATAAAGCTGTAACCTTATTCGTATAATTGGA 5384

```

```

DB 361  TTTGAGGCTTTAGACTTATAATTTCTTTGAATAAAACGTGATAACTTATTGTATAATTGGA 420
QY 5385  GTGGAGACCTACCTCCATAATTAGATAAACTCTTTTGGATTATATAATCAGAAATTTTGCT 5444
DB 421  GTGGAGACCTACCTCCATAATTAGATAAACTCTTTTGGATTATATAATCAGAAATTTTGCT 480
QY 5445  TTTTCTTCTCAAATTTTACATATGTATGTATATATATATATATCCACATATATAGTTTCCCT 5504
DB 481  TTTTCTTCTCAAATTTTACATATGTATGTATATATATATATATCCACATATATAGTTTCCCT 540
QY 5505  GATTAAATCGATATAAATAATATTCGGGTGCTTCAGGACTTTTGTCTTATATATTAAAG 5564
DB 541  GATTAAATCGATATAAATAATATTCGGGTGCTTCAGGACTTTTGTCTTATATATTAAAG 600
QY 5565  TATATTGTTTTATAGCAAGACATATTTCTGAATGTTTTTATAAACTTTTAAATAATTATA 5624
DB 601  TATATTGTTTTATAGCAAGACATATTTCTGAATGTTTTTATAAACTTTTAAATAATTATA 660
QY 5625  TGTAGTAATATTTTGTATCACAATGATATTTTTTTCTCTCTCTTCCCTTCCAAACTA 5684
DB 661  TGTAGTAATATTTTGTATCACAATGATATTTTTTTCTCTCTCTTCCCTTCCAAACTA 719
QY 5685  TACCACTG 5692
DB 720  TACCACTG 727

RESULT 11
LOCUS    AL596824
DEFINITION 717 bp mRNA linear EST 04-SEP-2003
            DKF2p451N191.r1.451 (synonym: hlcc1) spinal cord Homo sapiens cDNA
            clone DKF2p451N191.5', mRNA sequence.
ACCESSION  AL596824
VERSION    AL596824.1 GI:15154520
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 717)
  AUTHORS   Poustka,A., Wellenreuther,R., Mewes,H.W., Weil,B. and Wiemann,S.).
  TITLE     EST (Poustka,A., Wellenreuther,R., Mewes,H.W., Weil,B. and Wiemann,S.)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by DKFZ (German Cancer Research Center,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No si sequence available.
            This clone (DKF2p451N191) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..717
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKF2p451N191"
                     /tissue_type="human spinal cord"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="451 (synonym: hlcc1) spinal cord"
                     /note="vector: pSport1; Site_1: NotI; Site_2: SalI"
ORIGIN
Query Match      10.4%; Score 687.2; DB 9; Length 717;
Best Local Similarity 99.6%; Pred. No. 3.5e-126;
Matches 689; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAW465 row: k column: 02
High quality sequence start: 24
High quality sequence stop: 620.

FEATURES

source

ORIGIN

Query Match	10.3%	Score 682.6;	DB 14;	Length 935;
Best Local Similarity	98.7%;	Pred. No. 2.8e-125;		
Matches 741;	Conservative	0;	Mismatches 4;	Indels 6; Gaps 5;
Qy	3303	GAGGACAGTTTGCCCCACATCTCTGGTTGCACTCGGAGAAAAGGCATCAGTGCCTGTAGTA	3362	
Db	33	GATGACAGTTGCCCCACATCTCTGGTTGCACTCGGAGAAAAGGCATCAGTGCCTGTAGTA	92	
Qy	3363	GAGAACATCCATCTCATGAGCAGATCAATTTTCCTTCTTGAAGGTGAAGCCTTTCATCCT	3422	
Db	93	GAAGAACATCCATCTCATGAGCAGATCAATTTTCCTTCTTGAAGGTGAAGCCTTTCATCCT	152	
Qy	3423	GTTCACATTTGTCCTAAATGCTAATCTACTCGTGAATGTCAAATTCATATTTTATTCCTCA	3482	
Db	153	GTTCACATTTGTCCTAAATGCTAATCTACTCGTGAATGTCAAATTCATATTTTATTCCTCA	212	
Qy	3483	GACAAATATGGTACTTTTCCAAACAATGGATTCGATGGCTTGGGACAGGCAGAAAATTATT	3542	
Db	213	GACAAATATGGTACTTTTCCAAACAATGGATTCGATGGCTTGGGACAGGCAGAAAATTATT	272	
Qy	3543	ATTCTATTGTATGTTTGGCCAAATGAAGATACTAATTCCTAAGGACATCTTTCAGACTATTT	3602	
Db	273	ATTCTATTGTATGTTTGGCCAAATGAAGATACTAATTCCTAAGGACATCTTTCAGACTATTT	332	
Qy	3603	ATCACCATAATATAAGGATGCTCTAAAAGGAAAAATACATAGAAAACCTTGGACAATATTACC	3662	
Db	333	ATCACCATAATATAAGGATGCTCTAAAAGGAAAAATACATAGAAAACCTTGGACAATATTACC	392	
Qy	3663	TTTACTCAGAGTTTCTCAGTAGCAAGGATCAGGAGGATTCCTGTGTTATTATACACCTACT	3722	
Db	393	TTTACTCAGAGTTTCTCAGTAGCAAGGATCAGGAGGATTCCTGTGTTATTATACACCTACT	452	
Qy	3723	TTTCAGAAACTTCATGATCTCTCATTACCAAGTAATCCTTTTCTTTGTGGAATCTTTATC	3782	
Db	453	TTTCAGAAACTTCATGATCTCTCATTACCAAGTAATCCTTTTCTTTGTGGAATCTTTATC	512	
Qy	3783	CAGNAGCTTCAGATTCCTCGGGCAAGGTTTTTCCCTATGCTTTTAATGTTTGAGATTGGGT	3842	
Db	513	CAGAAGCTTCAGATTCCTCGGGCAAGGTTTTTCCCTATGCTTTTAATGTTTGAGATTGGGT	572	
Qy	3843	GCAGAAATATAAGCATATCTGCTCTCTAAAGCATACAG-AGGCCGAAAAACCTCTTTT	3901	
Db	573	GCAGAAATATAAGCATATCTGCTCTCTAAAGCATACAGAGGCCGAAAAACCTCTTTT	632	
Qy	3902	TGGAGAAATAGGACACACTAATGACTTACCTTG-TTGAACCTTCGAAATTACCAGTATA	3960	
Db	633	TGGAGAAATAGGACACACTAATGACTTACCTTGTTTGTGACCTTCGAAATTACCAGTATA	692	
Qy	3961	CCTTGCATAATATAGATCAACT-GTTGATTCAATATGG--AAATGGGAAAAAGCTGCATAA	4017	

QY	628	GTGATCTCATTAAGTGTGACATGGGTAACTTCTAGTTCATGCAACCAATAGTGAAGAGATATTA	687
DB	26	GGGATCTCATTAAGTGTGACATGGGTAACTTCTAGTTCATGCAACCAATAGTGAAGAGATATTA	85
QY	688	AAAAATTAATGCCAGATGATTTTAAAGTCTAAATCGAGATTCCTTGATTGGATTGGATTTAT	747
DB	86	AAAAATTAATGCCAGATGATTTTAAAGTCTAAATCGAGATTCCTTGATTGGATTGGATTTAT	145
QY	748	CTTCAGTGTCCAGATACCTCCCTGTGTTCTTCTCAACAGACCATGATAGTGTGATCTGTCCAG	807
DB	146	CTTCAGTGTCCAGATACCTCCCTGTGTTCTTCTCAACAGACCATGATAGTGTGATCTGTCCAG	205
QY	808	AACACACGAATGATATCAGTTCCTGAATTCACAAATACAGAAATCGGAGGAATCAAGAAT	867
DB	206	AACACACGAATGATACCAAGTTCCTGAATTCACAAATACAGAAATCGGAGGAATCAAGAAT	265
QY	868	TGGGTATAAAAAGTAGATACACACTTTTCAGATTCCTATTAATTACAGTGGAAACAGAAAAAT	927
DB	266	TGGGTATAAAAAGTAGATACACACTTTTCAGATTCCTATTAATTACAGTGGAAACAGAAAAAT	325
QY	928	TAAAGATATAAAGATCTTTAATCAGTTAGAAATCAATTTGTTGATTTTACATGTCTCATCTG	987
DB	326	TAAAGATATAAAGATCTTTAATCAGTTAGAAATCAATTTGTTGATTTTACATGTCTCATCTG	385
QY	988	CTTTCACTCGACAAAGTTCCTAAAATGTTTCATGCCAAAGACAAGCTACAAACAAGAGGCC	1047
DB	386	CTTTCACTCGACAAAGTTCCTAAAATGTTTCATGCCAAAGACAAGCTACAAACAAGAGGCC	445
QY	1048	AGCCATGTGGATTAATAAGATGTTGGCTTAGTATAAAGAGGAAGTAGATGTGGCAGTCA	1107
DB	446	AGCCATGTGGATTAATAAGATGTTGGCTTAGTATAAAGAGGAAGTAGATGTGGCAGTCA	505
QY	1108	TAACTGCCGCGAATGTTTAAAGACAGAGGCGAAGACAAGTGCTTTGACCTCGACGCTTC	1167
DB	506	TAACTGCCGCGAATGTTTAAAGACAGAGGCGAAGACAAGTGCTTTGACCTCGACGCTTC	565
QY	1168	CGAAAAATGAAGATTTATGCTTTAAATGATTCAATTCAGAGATGAAAAATTTCAAATTAC	1227
DB	566	CGAAAAATGAAGATTTATGCTTTAAATGATTCAATTCAGAGATGAAAAATTTCAAATTAC	625
QY	1228	CTGACTTTTCTTTCAGGAAGATTAAGACTGTTATTAACCAATCTGCACAGGAAGACTCAA	1287
DB	626	CTGACTTTTCTTTCAGGAAGATTAAGACTGTTATTAACCAATCTGCACAGGAAGACTCAA	685
QY	1288	AAAGTTTATAGACCTTAAGGATTAATGATGAATC	1319
DB	686	AAAGTTTATAGACCTTAAGGATTAATGATGAATC	717

1288 AAAGTTTAGACCTTAAGGATAATGATGTAATC 1319
|||||
686 AAAGTTTAGACCTTAAGGATAATGATGTAATC 717

RESULT 12	CD557100	935 bp	mRNA	linear	EST 11-JUN-2003
LOCUS	CD557100				
DEFINITION	AGENCEOT 14413152 NIH MGC 180 Homo sapiens CDNA clone				
	IMAGE:30350481	5'	mRNA	sequence.	

ORGANISM	<p>Homo sapiens</p> <p>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</p>
REFERENCE	<p>1 (bases 1 to 935)</p>
AUTHORS	<p>NIH-MGC http://mgc.nci.nih.gov/.</p>
TITLE	<p>National Institutes of Health, Mammalian Gene Collection (MGC)</p>
JOURNAL	<p>Unpublished (1999)</p>
COMMENT	<p>Contact: Daniela S. Gerhard, Ph.D.</p> <p>Office of Cancer Genomics</p> <p>National Cancer Institute / NIH</p> <p>Bldg. 31 Rm10A07 Bethesda, MD 20892</p> <p>Email: cgabs@email.nih.gov</p> <p>Tissue Procurement: Dr. Michael Brownstein</p> <p>CDNA Library Preparation: Invitrogen Corp</p>


```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="IMAGE:30318937"
/note="Vector: pBluescriptSKm"

ORIGIN
Query Match      10.2%; Score 676.4; DB 10; Length 711;
Best Local Similarity 97.7%; Pred. No. 4.9e-124;
Matches 586; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 4842 TTTTTCATTATAGAACATCTTTTATAGTGAAGAATGCGCATATACATATGCAACCT 4901
DB 702 TTTTCCATAAAAAAATCTTTTATAGTGAAGAATGCGCATATATCATATGCAACCT 643

QY 4902 AATTGTTAAACTAACTCCAGCCTAAAGCTGAAATGCCACAACACTAAAGTATATAA 4961
DB 642 ATTTTGTAAACTAACTCCAGCCTAAAGTGAATGCCCTCAACACTAAAGTATATAA 583

QY 4962 TATGTCGTGATTTTGAACACATAGCTTTGCTCTTTAGGCGAGGAATGATCTTTTCAAAT 5021
DB 582 TATGTCGTGATTTTGAACACATAGCTTTGCTCTTTAGGCGAGGAATGATCTTTTCAAAT 523

QY 5022 CATTAGCACAATTTTAATATCTTAAATTTAAGATTCATATCTTCTGTAGCTTTTAC 5081
DB 522 CATTAGCACAATTTTAATATCTTAAATTTAAGATTCATATCTTCTGTAGCTTTTAC 463

QY 5082 AATTAAATTTAAGTACTAAAAAGACAAGGATTTCTTTTAAAGAAATTTATAGCAATTTACTGT 5141
DB 462 AATTAAATTTAAGTCTTAAAGACAAGGATTTCTTTTAAAGAAATTTATAGCAATTTACTGT 403

QY 5142 GTTATTTAAATGCTTAAGCCAAAGTATCTGCACTTAGGTATACCTTTTATGCCAATTAATG 5201
DB 402 GTTATTTAAATGCTTAAGCCAAAGTATCTGCACTTAGGTATACCTTTTATGCCAATTAATG 343

QY 5202 ATTTTAAATGAAGGCTCTTTTCAGATGTAACCTTTATGAAGAAATATCTGCTTTGTGATA 5261
DB 342 ATTTTAAATGAAGGCTCTTTTCAGATGTAACCTTTATGAAGAAATATCTGCTTTGTGATA 283

QY 5262 TGCCAGTTTGAATACTGTTTCTTAAAGTCTGCTCAAAATGTTATTTTCAAGTGCACAAAAACC 5321
DB 282 TGCCAGTTTGAATACTGTTTCTTAAAGTCTGCTCAAAATGTTATTTTCAAGTGCACAAAAACC 223

QY 5322 AGTTTGAAGTCTTAGACTTATATCTTTTCAATTAAGAACTGATTAATTTTGTATTAAT 5381
DB 222 AGTTTGAAGTCTTAGACTTATATCTTTTCAATTAAGAACTGATTAATTTTGTATTAAT 163

QY 5382 GGAGTGGAGACCTACCTCCATAATTAGATAAACTCTTTTGGATTATAATCAGAAATTTTG 5441
DB 162 GGAGTGGAGACCTACCTCCATAATTAGATAAACTCTTTTGGATTATAATCAGAAATTTTG 103

QY 5442 CTTTTTCTTCTGAAATTTACATATGATATGATATATATATATCCACATATATAGTTTC 5501
DB 102 CTTTTTCTTCTGAAATTTACATATGATATGATATATATATATCCACATATATAGTTTC 43

QY 5502 CCGATTAAATGGATATTAATAATTAATGCGGGTCTCTCAGGA 5543
DB 42 CCGATTAAATGGATATTAATAATTAATGCGGGTCTCTCAGGA 1

RESULT 15
CB962678/c 758 bp mRNA linear EST 29-APR-2003
LOCUS
DEFINITION AGNCOURT 13465015 NIH MGC 187 Homo sapiens cDNA clone
IMAGE:30318937 5', mRNA sequence.
ACCESSION CB962678
VERSION CB962678.1 GI:30218796
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 758)

```

AUTHORS TITLE JOURNAL COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapps-r@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDCM114 row: f column: 02
High quality sequence stop: 611.

FEATURES

source

1. 758
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30318937"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 187"
/note="Organ: Blood_vessels - aorta, basilar and artery;
Vector: pDNR-LIB; Site 1: SfiI (ggccattatggc); Site 2:
SfiI (ggccgctcggc); 5' and 3' adaptors were used in
cloning as follows: 5' adaptor sequence:
5'-ATTCAGAGCGCGAGCGCGACATG-DT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.4 kb
(range 0.5-4.0 kb). 14/15 colonies contained inserts by
PCR. This library was enriched for full-length clones and
was constructed by Clontech Laboratories (Palo Alto, CA).
Note: this is a NIH_MGC Library."

ORIGIN

Query Match 10.1%; Score 668.4; DB 14; Length 758;
Best Local Similarity 99.7%; Pred. No. 1.9e-122;
Matches 669; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 679 AGAGATATTAATAAATATTTGCGAGATGATTTTAAGTCTTAATGCGAGATTCCTTGATGAT 738
DB 672 AGAGATATTAATAAATATTTGCGAGATGATTTTAAGTCTTAATGCGAGATTCCTTGATGAT 613

QY 739 TGGATTTATCTTCAGTGTGAGATCTCCCTGTGTTCTTCAACGACCATGATAGTATA 798
DB 612 TGGATTTATCTTCAGTGTGAGATCTCCCTGTGTTCTTCAACGACCATGATAGTATA 553

QY 799 CTGTCAGAGAACACAGAAATGATATCAGTCTTGAATTAATAAATAGAGAAATCGAGGAA 858
DB 552 CTGTCAGAGAACACAGAAATGATATCAGTCTTGAATTAATAAATAGAGAAATCGAGGAA 493

QY 859 TCAAGAAATGGGTAAAGTAGATACAACATTTTCAGATTCCTATTAATACAGTGA 918
DB 492 TCAAGAAATGGGTAAAGTAGATACAACATTTTCAGATTCCTATTAATACAGTGA 433

QY 919 CAGAAATTTTAAAGATATAAAGATCTTTAATCAGTTAGAATCAATTTGATTTTACA 978
DB 432 CAGAAATTTTAAAGATATAAAGATCTTTAATCAGTTAGAATCAATTTGATTTTACA 373

QY 979 TGTCTATCTGCTTTGACTGACAAAGTTCAAAATGTTTCNTGCCAACACAGCTACAC 1038
DB 372 TGTCTATCTGCTTTGACTGACAAAGTTCAAAATGTTTCNTGCCAACACAGCTACAC 313

QY 1039 ACAAGAGCCAGCATCTGGATTACTATAAAGATGTTGGCTTAGTAAAGAGGAAGTAGATG 1098
DB 312 ACAAGAGCCAGCATCTGGATTACTATAAAGATGTTGGCTTAGTAAAGAGGAAGTAGATG 253

QY 1099 TGGCAGTCATAACTCCCGCAGAAATGTTTAAAGAGAGGCGCAAGCAAGTCTTTGACCT 1158
DB 252 TGGCAGTCATAACTCCCGCAGAAATGTTTAAAGAGAGGCGCAAGCAAGTCTTTGACCT 193

QY 1159 CGAGCCTTCCGAAAATGAAGATTTATCTTAAATGATTTCAAAATTCAGAGATGAAAT 1218

```
Db      192 GCAGCCTTCGAAAAATGAAGATTATGCTTAAATGATTCAAATTCAGAGATGAAAAATT 133
Qy      1219 TCAAAATFACCTGACTTTTCCTTTCAGGAGAGATAAGACTGTATATAAACAATCTGCACAAG 1278
        |||||
Db      132  TCAAAATFACCTGACTTTTCCTTTCAGGAGAGATAAGACTGTATATAAACAATCTGCACAAG 73
        |||||
Qy      1279 AAGACTCAAAAAGTTTAGACCTTAAGGATAATGATGTAATCCCAAGATTCTCTTCAGCTT 1338
        |||||
Db      72  AAGACTCAAAAAGTTTAGACCTTAAGGATAATGATGTAATCCCAAGATTCTCTTCAGCTT 13
        |||||
Qy      1339 TACATGTTTCC 1349
        |||||
Db      12  TACATGTTTCC 2
```

Search completed: May 18, 2004, 07:03:30
Job time : 10173.9 Secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 16:37:02 ; Search time 290.233 Seconds
(without alignments)

12680.960 Million cell updates/sec

Title: US-09-744-167-3

Perfect score: 6632

Sequence: 1 actccggccgggtagctc.....ttggaggcacatttgaagt 6632

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgm2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgm2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgm2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgm2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgm2_6/ptodata/2/ina/PTCUS_COMB.seq.*
6: /cgm2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	64.4	1.0	13124	2	US-08-487-826B-13
2	62.8	0.9	8961	4	US-10-204-708-80
3	61	0.9	6152	3	US-08-973-462-1
4	59	0.9	13124	2	US-08-487-826B-13
5	57.8	0.9	5666	4	US-10-204-708-30
6	57.4	0.9	6156	4	US-10-204-708-59
7	56.8	0.9	7218	1	US-08-232-433-14
8	56.8	0.9	53332	4	US-09-801-861-3
9	56.4	0.9	640681	4	US-09-790-988-1
10	56	0.8	6124	4	US-08-213-419B-3
11	55.2	0.8	636	3	US-08-998-416-1137
12	55.2	0.8	837	3	US-08-998-416-288
13	53.8	0.8	6113	4	US-10-204-708-14
14	53.6	0.8	6156	4	US-10-204-708-60
15	53	0.8	832	4	US-09-521-976-2813
16	52.8	0.8	640681	4	US-09-790-988-1
17	52.6	0.8	665	2	US-08-883-795A-36
18	52	0.8	615	3	US-08-998-416-186
19	52	0.8	6866	4	US-10-204-708-19
20	51.6	0.8	2095	6	5231168-1
21	51.4	0.8	5852	1	US-07-867-106-2
22	51.2	0.8	658	3	US-08-998-416-595
23	51.2	0.8	5340	4	US-09-627-122-21
24	51	0.8	20674	4	US-09-641-638-651
25	50.4	0.8	832	4	US-09-621-976-2813
26	50.4	0.8	6243	2	US-09-056-075-1
27	50.2	0.8	6801	4	US-10-204-708-61

28	49.6	0.7	6113	4	US-10-204-708-13	Sequence 13, Appl
29	49.4	0.7	11049	4	US-10-204-708-22	Sequence 22, Appl
30	49	0.7	1183	2	US-08-731-722-8	Sequence 8, Appl
31	49	0.7	4140	3	US-08-894-731-2	Sequence 2, Appl
32	49	0.7	6317	4	US-10-204-708-11	Sequence 11, Appl
33	48.8	0.7	10467	4	US-10-204-708-2	Sequence 2, Appl
34	48.6	0.7	5152	4	US-10-204-708-73	Sequence 73, Appl
35	48.6	0.7	168575	4	US-09-426-290-1	Sequence 1, Appl
c	48.4	0.7	724	3	US-08-998-416-683	Sequence 683, App
36	48.4	0.7	6070	4	US-10-204-708-10	Sequence 57, Appl
37	48.4	0.7	6326	4	US-10-204-708-57	Sequence 472, App
38	48.2	0.7	767	3	US-08-998-416-472	Sequence 8, Appl
39	48.2	0.7	1511	1	US-07-991-867B-8	Sequence 8, Appl
40	48.2	0.7	1511	1	US-08-107-755A-8	Sequence 8, Appl
c	48.2	0.7	1511	2	US-08-544-332-8	Sequence 8, Appl
41	48.2	0.7	1511	4	US-09-370-861A-8	Sequence 8, Appl
c	48.2	0.7	1511	4	US-08-107-755A-1	Sequence 1, Appl
c	48.2	0.7	8457	1	US-07-991-867B-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-08-487-826B-13

; Sequence 13, Application US/08487826B

; Patent No. 593827

; GENERAL INFORMATION:

; APPLICANT: Sim, Kim L.

; APPLICANT: Chitnis, Chetan

; APPLICANT: Miller, Louis H.

; APPLICANT: Peterson, David S.

; APPLICANT: Su, Xin-zhaun

; APPLICANT: Wellens, Thomas E.

; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

; NUMBER OF SEQUENCES: 45

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobbe Martens Olson & Bear

; STREET: 620 Newport Center Drive 16th Floor

; CITY: Newport Beach

; STATE: California

; COUNTRY: US

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/487,826B

; FILING DATE: 10-SEP-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Israel, Ned

; REGISTRATION NUMBER: 29,655

; REFERENCE/DOCKET NUMBER: NIH121.001CPI

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 235-8550

; TELEFAX: (619) 235-0176

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 19124 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; US-08-487-826B-13

Query Match 1.0%; Score 64.4; DB 2; Length 19124;
Best Local Similarity 45.3%; Pred. No. 6.4e-05;

QY 5790 TCACATTAATCCTAGAAATTTCTTGGCAACTTCGCATATTTTCATTCGACACTGGTGTGA 5849
Db 6103 TTTTTCGTTTTCGTTTATTTGTAATATTTTAAATTTTTCGTTTAAAGTAGTATT 6162
QY 5850 TAAGTATAAATTTAAAGAACTAATTAATTTTCATATTTTAAATTTCTTTATAT 5903
Db 6163 TTGTAAAGAGTTTTCGATGTTGGGATTTAGTTTTCATATTTTATTTATTT 6216

RESULT 3

US-08-973-462-1
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRULHE, PIERRE
; APPLICANT: DUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6152
; TYPE: DNA
; ORGANISM: P. falciparum
US-08-973-462-1

Query Match 0.9%; Score 61; DB 3; Length 6152;
Best Local Similarity 48.7%; Pred. No. 0.00025;
Matches 166; Conservative 0; Mismatches 175; Indels 0; Gaps 0;
QY 5336 AGACTTAATCTTTGAAATAAAGTCAATACATTTATTTGTAATTCGAGTGGAGACCTA 5395
Db 5716 AT 5775
QY 5396 CCTCCAAATPAGATAAACTCTTTTGGATTATATATATATATATATATATATATAT 5455
Db 5776 TATTAAATTTTATATTTTCCATATATATATATATATATATATATATATATATATAT 5835
QY 5456 AAATTTATACAT 5515
Db 5836 AATGTTTTTACAGATTTATGTTTTTTTAAATTAATATATATATATATATATATAT 5895
QY 5516 TATTAAATATATTCGGGGTCTTCAGGACTTTTTCCTCTATATATTTAAGTATATGTTT 5575
Db 5896 TATTATTCATAGAT 5955
QY 5576 TATAGCAAGACATATCTGAATGTTTTTATATATATATATATATATATATATATATAT 5635
Db 5956 TAAT 6015
QY 5636 TTTTGTATCAATGCAATTTATTTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5676
Db 6016 TTGTATGCAATTAAGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 6056

RESULT 4

US-08-487-826B-13/C
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.

; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
US-08-487-826B-13

Query Match 0.9%; Score 59; DB 2; Length 19124;
Best Local Similarity 43.0%; Pred. No. 0.0011;
Matches 504; Conservative 0; Mismatches 660; Indels 9; Gaps 4;
QY 5035 TTTAAATATCTAAATTTAAGAGATCCATATCTTCTGTAGCTTTACATTAATTTAAGT 5094
Db 7062 TATTATATATACAT 7003
QY 5095 ACTAAAAGACAGGATTTCTTTTAAAGAAATTTATAGCAATTTACTGTGTTTAAATGC 5154
Db 7002 ATATATTTAGTTTATATATATTTTAAAGAAATATATATATATATATATATATAT 6943
QY 5155 TAAGCCAAAGTATCTGCATCTAGGTATACCTCTTTATGCCAATAATGATTTTAAAGAG 5214
Db 6942 TAAATTTAAATATCTTACAAAAAATAATATATATATATATATATATATATATAT 6883
QY 5215 CTCTTTTCAGATGAACCTTTATGAAGGAATATCTCTTTGTGTATATCCAGTTAGAT 5274
Db 6882 CTTATTTATCTATTTAATTTATATATATATATATATATATATATATATATATAT 6823
QY 5275 ACTGTTTCTAAAGTCTGCAAAATTTGTTTTCAGTGGCAAAAAACCAAGTTTTCAGGCT 5334
Db 6822 AATTAATGAGAAAAAATAATGAAAAATACAAACATATAAAAAAGTATATATATGCA 6763
QY 5335 TAGACTTATAATTTCTTTGAATAAACTGATACTTTATTTGTATATTTGAGTGGAG ---A 5391
Db 6762 CGTGTATATATATTTAATTTAATTAATATATATATATATATATATATATATATAT 6703
QY 5392 CCTACCTCCATATATAGATAAATCTTTTGGATTATATATATATATATATATATATAT 5451
Db 6702 ATTAT 6643
QY 5452 TCTCAAAATTTATCAT 5511
Db 6642 TCTATCTATATATATCTAT 6583

QY 5512 TCGATATTAAATAATTCGGGTGCTTCAGGACTTTTTCCTCTATATTAAAGTATATTG 5571
Db 6582 ATATCTTTTGTATGTTGTTATTAATAAGTAGATATTATTAATTATTTTCTT 6523
QY 5572 TTTTATAGCAAGACATATCTGAAATCTTTTATAAATCTTTAATAATTATATGAGGT 5631
Db 6522 GTTATATATTAAATTTATTAATAATAATATATAGTACCAAAAAAATAATGATTAAT 6463
QY 5632 AATAATTTTGTATCAATGCAATTTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5691
Db 6462 AATAACAGGAAAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6403
QY 5692 GTATTACCACTTCAAGAGTACTGACGACGGCCAGATGACCCCTGGAAGTAGTCATTA 5751
Db 6402 TCATTGTG--TTTGTGTATTTTATGTTATTCATGCAITTTATGAATTTCAAAATTT 6345
QY 5752 TGTAGCAATAATGAAGCTGAAACAGGTTTCTTCTCTCTCTCTCTCTCTCTCTCTCT 5811
Db 6344 TATTGTATAATAATAAAAAAATAAGTAAAAATACACATTTATAAATATATATTTCAAAATAT 6285
QY 5812 TCTTGGCAACTTCGCATATTTTCATGACACTGGTGTATAGTATAAATTTAAATGAAC 5871
Db 6284 GAGTTATTAAATAAATGTTCACTTCTATATTTTATATAAAGAAATATTTGTTA--- 6228
QY 5872 AATTAATTTTGCATATTTTAAATCTTTTATATGAGTATTTTATTAACAGGATATTA 5931
Db 6227 TAAATAAATAATATATGCTACTATATAAATAATTAATAATCTTTAAAGTATATACTA 6168
QY 5932 ACATAAGTTAAATCTATGTTTGAATTTGAAATTTGAAAGCTTTCTCTCTCTCTCTCT 5991
Db 6167 AAATATAAATAATGCACTGTA--TAAATAAGTATAAATCAATCATATATATATATAT 6109
QY 5992 GCAAAAAAGTGGGGGCATATTGTAGTCTCTGTCATTTAAGTTATGTAAAAAATTAATCA 6051
Db 6108 AT 6049
QY 6052 TTATTTTCTGATCTTTAAACATCTCATGTGTAATATATATGTTTGTATCAAAACATCA 6111
Db 6048 TATCCAAATCTGATATGTTTATATTTTATTTATTTATTAATAACAAAAAAGAACACA 5989
QY 6112 TATATTCAAGAAAAAATAATGTTAAATAGCCCTGTTTAAAGAAAAATATTTATGAA 6171
Db 5988 AGAGTAAACAAAAAATAATTAATAATAAAGATGATATAATAAATAATTAATTAATCA 5929
QY 6172 GCATCTCACTTGAAGTCAAGTCAAGTTATA 6204
Db 5928 TTAGTATAGTTTATGTTCTTTATATATATATA 5896

RESULT 5
US-10-204-708-30
; Sequence 30, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: US/10204708
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98

; SEQ ID NO 30
; LENGTH: 5666
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-30

Query Match 0.9%; Score 57.8; DB 4; Length 5666;
Best Local Similarity 45.7%; Pred. No. 0.0013;
Matches 293; Conservative 0; Mismatches 332; Indels 16; Gaps 2;

QY 5067 TTTCTCTAGCTTTACAAATTAATTTAAAGTACTAAAAAGACAGGATTTCTTTTAAAGAAAT 5126
Db 3308 TTGGGAATATTAAAGATTTGTTAGGAGTTTGTAGGTTAGAAATTTTATATAATAT 3367
QY 5127 TATAGCATTTACTGCTTATTTTAAATGCTAAAGCAAGTATCTGCATCTAGGTATACCTC 5186
Db 3368 TTAATAATTTGATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 3427
QY 5187 TTTATGCCAATATGATTTTAAATGAAGCTCTTTTTCAGATGTAACCTTATGAAGAAAT- 5245
Db 3428 ATTATATGATATATGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 3487
QY 5246 -----ATCTGCTTTGCTATATGCCAGTTAGAATPACTGGTTTCTTAAAGTCTGCA 5295
Db 3488 GTTTTATTAAAGTTGATAGTAGTGAAGCAATTTGTAATAATTTAGTAATTTATTTTATTT 3547
QY 5296 AATTGTTATTTTCAGTGGCACAACCAAGTTTGGAGTCTTAGACTTATAATTTCTTTGAAT 5355
Db 3548 TATTTTCTTTTGTGTTTAAAGATATAGTTATGTTTATATATATGTTTATTTATTTATTA 3607
QY 5356 AAAAAGTGAATCTTATTTGTTATAATTTGGAGTGGAGACCTACCTCCATTAATTTAGATA 5415
Db 3608 ATAATGGGTTAATTTATTTTATTTTAAAT-----AAATTAATAAGTAAATAGGTTAAAT 3662
QY 5416 CTTTITGGATTAATATCAAGATTTTGCCTTTTCTCTCAATTTATACATATGATGT 5475
Db 3663 ATATTAGTTTAAATTTTAAATAATAAATAATTAATAGTTAAATTTTATATATATTT 3722
QY 5476 ATTATATATCCATATATAGTTTCCCTGATTAAATGATTAATAATAATTTTGGGGTG 5535
Db 3723 TTTTGAAGTTTATAGTATTTTAAAGTGTAAAGGGTTTGTAAATTTAAATTTGGAG 3782
QY 5536 CTTTCAAGCTTTTCTCTATATTTTAAAGTATATTTTATTTTATCAAGAACATATTTCTG 5595
Db 3783 GTTTGTGAGTTATAGTAAATTTAGAGAGTTAAAGTGATTTTCTTTATATATAGTATG 3842
QY 5596 AATGTTTATAAATCTTTAATAATTTTATATGATGTAATATTTTGTATCAATGCAAT 5655
Db 3843 TGTATAGATAGTTTGTATTTTATTTTGTATTTTATGCGTTTATTTTGTTTTATTTG 3902
QY 5656 ATTTTCT 5696
Db 3903 TGTATTAGTTTGTATTTTATTTTATATATATATATATATATATATATATATATAT 3943

RESULT 6
US-10-204-708-59
; Sequence 59, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: US/10204708
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8

```

, PRIOR FILING DATE: 2000-04-06
, PRIOR APPLICATION NUMBER: DE 10019173.8
, PRIOR FILING DATE: 2000-04-07
, PRIOR APPLICATION NUMBER: DE 10032529.7
, PRIOR FILING DATE: 2000-06-30
, PRIOR APPLICATION NUMBER: DE 10043826.1
, PRIOR FILING DATE: 2000-09-01
, NUMBER OF SEQ ID NOS: 98
, SEQ ID NO 59
, LENGTH: 6156
, TYPE: DNA
, ORGANISM: Artificial Sequence
, FEATURE:
, OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-59

```

Query Match	0.93;	Score 57.4;	DB 4;	Length 6156;
Best Local Similarity	43.3%;	Pred. No. 0.0017;		
Matches 318; Conservative	0;	Mismatches 416;	Indels	1; Gaps 1;
QY	5401	ATAATTAGATAAACTCTTTTGGATTATAATCAGAAATTTTGCCTTTTCTTCTCTCAAAAT	5460	
DB	1193	ATAATTTGTAATAATATTAGTATATATGTCGGGTATTGATGTTGTTTTTAAGTA	1252	
QY	5461	ATTACATATGATGTATTATATATCCACATATATAGTTTTCCCTGATAAATGGAATTA	5520	
DB	1253	ATATAAATTAAAGAAATTAATTTTTTAAATTTATATTAAAGTTTTTTTTTAGTGATTTTA	1312	
QY	5521	AAATAATTCGGGTGCTTCAGGACCTTTTTCCTCTATATTAAAGTATATTGTTTTTATAG	5580	
DB	1313	CGTAAAGGTATTCGGGGATAAAATTTTGGAGAAAAAAATCGTTGATATAAATTTTAAAT	1372	
QY	5581	CAAGAACATATTCCTGAATGTTTTTATAAATCTTTTAAATTTATATGTCAGGTAATTTTT	5640	
DB	1373	ATTTTTAGTATATAAAGAGAGATTATTTTGATTTTGAATTTAATGTTAGTTAGTTTTAT	1432	
QY	5641	GATACAAATGCATATTTTTTTTTCCTCTTTTCCTCCAAACTATACCACTGATTTTACC	5700	
DB	1433	TAATGACAGTCAVATTTTGGGGTTTGAATTTAGTGTTTTGGTTTTTAAATATGATTTAGATG	1492	
QY	5701	ACTTCTAAGAGTGACTCAGCAGCGGCCAGATGACCCCTTGAAGTAGTCATTATTGTAGCAAT	5760	
DB	1493	GAITTTATTGTGTGGAGGTGATTAAGATTAAAGATAAAGATAATAATTTATTATTAGAAGA	1552	
QY	5761	AAATGAAGCCCTGAACACAGGTTTTTTTTTACTTCCACTTTTAAATCCTTTAGAAATTTCTTGCGAA	5820	
DB	1553	AAAATATAGTTCAATAGATGTAATTTTAGTTAAATTAATAGTAGCGGTTCGTATATATAT	1612	
QY	5821	CTTCGCATATTTCCATTGCACACTGGTGTATPAAGTATAAAATTTAAATGAACATACTTT	5880	
DB	1613	AAATATTATATTATTATTATTTAAAAAAATTTGTTAGGTGTGAAAAAATTTGTTGTGATTTTT	1672	
QY	5981	TGCATATTTTAAATCTTTATATGATGTTATTTTTTATACAGGATATTAAACATAAGTT	5940	
DB	1673	TTGAAAGATAATTTTAAATATAGGTAAATGGGTGAGTTTGTGTAAAGTATAGAAGATA	1732	
QY	5941	AAATCCTATGPTATTGAAATTTGTACAGAGCTTTCCCTCTTACCTCAACACAGCAAAAAG	6000	
DB	1733	TTTTTGGAAATGTTTGAATTTTTTTTTTAAATAT-ATTATATATTAATTTTTTAAATATA	1791	
QY	6001	TGGGGGGCATATTTGTAGTCCTGTCATTTAAGCTTATGTAABAAATTTAATCATTTATTGA	6060	
DB	1792	TTTTTTTTTGTAGTTGGAATTTTGTATTTTAAATTTATTTTTTGAAGAAAAAAATTTTATTT	1851	
QY	6061	TGCTTTAAACATCTCTGTGTATAATATGTTTTTGTATCAAAACACTCATATATTTC	6120	
DB	1852	TTTTTAGAATTTAAATATTTTGTAAAAATAGGATTTATTTTTCGAAGTGTGTATCGTTTT	1911	
QY	6121	AGAAAAAGAAATTAT	6135	
DB	1912	AAGGTAGTAATTGT	1926	

RESULT 7
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: ptz9pt-Fls
US-08-232-463-14

[illegible]

NAME/KEY: CDS
LOCATION: (2407) .. (2439)
NAME/KEY: CDS
LOCATION: (2598) .. (3404)
NAME/KEY: CDS
LOCATION: (3580) .. (3720)
NAME/KEY: CDS
LOCATION: (3850) .. (5835)
US-08-213-419B-3

Query Match 0.8%; Score 56; DB 4; Length 6124;
Best Local Similarity 46.8%; Pred. No. 0.0036;
Matches 289; Conservative 0; Mismatches 315; Indels 14; Gaps 3;

Qy 5546 TTTTGGCTTCTATATTAAGTATATGTTTATAGCAAGAACATATCTCAATGTTTAT 5605
Db TTTCTCTGACATTTTATACATTTTGTGTAATATACACCTAAATAAATATGAC 2581

Qy 5606 AAATCTTTAAATTAATATATAGTAAATATTTTATGATCAATGCAATTTTTC 5665
Db AAATGTTTGTGTAATGCGTAAATATTTATTTATATATATATATATATTTTAC 2521

Qy 5666 CTCCTTCTTCCAACTATACCACTGATTTACCACTTTAGAGTACTGACGCGG 5725
Db ATATTTGTTTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 2463

Qy 5726 CCAGATGACCTTGAAGTAGTATATGATAGCAATAAATGAAGCTCGAAACAGGTTTT 5785
Db TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 2403

Qy 5786 TACTTCCACTTAATCTTGAATATTTCTTGGCACTTCGCAATTTTCATTTGACACTGG 5845
Db ATATATGAATATTAAGTATATATATATTTTAAAGAACTT---TAAATTTTTCGAATGT 2347

Qy 5846 TGTATAGTATAATTTAAATGAATTAATTTCTTTGCAATTTTAAATCTTTTATATGG 5905
Db TGTATTTTAAATTTCTTTATGTTTAAATATATATATATATATATATATATAT 2287

Qy 5906 TAGTATTTTATACAGGATATTAACATAAGTTAAATCTATGTTTAAATTTGAAATGTTA 5965
Db TTAATTTTATTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 2227

Qy 5966 CAGACTTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 6025
Db TAT 2175

Qy 6026 TTTAAGTTAATGTAATAAATTTAATCAATTTATTTGATGCTTTTAAACATTTCTCATGTAAT 6085
Db TAAATTTTATTTTATTTTAAATTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 2115

Qy 6086 ATATGTTTGTATCAAAACACTCATATATTTCAAGAAAGAAATATGTTTAAATAGC 6145
Db TAAATCTAGATAATGACCTCTTTTATTTTAAATTTTCAAAAAAATATATATGACACC 2055

Qy 6146 CCTGTTTAAAGAAATA 6163
Db TCTCAATAAATAAATAA 2037

RESULT 11
US-08-998-416-1137
Sequence 1137, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Reischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPHII
TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998.416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PE/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8567
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1137:
SEQUENCE CHARACTERISTICS:
LENGTH: 636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1692RP
US-08-998-416-1137

Query Match 0.8%; Score 55.2; DB 3; Length 636;
Best Local Similarity 49.2%; Pred. No. 0.0022;
Matches 174; Conservative 0; Mismatches 178; Indels 2; Gaps 1;

Qy 5327 TGAGGCTTAGACTTATAATCTTTGAATAAAACGATAAATTTTGTATTAATTTGGAGT 5386
Db TTAATTTTATTTGATTTAATTAATCTTATATATATATATATATATTAATTTACTTAATTCATCA 164

Qy 5387 GGAGACCTACCTCCATTAATAGATAAATCTTTTGGATTAATACAGAAATTTGCCTTT 5446
Db TTATTAATATTTTATATAAATATAAATAATATTTTAAATGAATACTATTTAGTCTATGT 224

Qy 5447 TTTCTTCTCAATTAATTAAT 5506
Db TCAATTTTAAATTTAGTTTAAAT 284

Qy 5507 TTAATATGATTAATAAATAATTTGCGGCTCTTCCAGACCTTTTGTCTTCTATATTTAAGTA 5566
Db TTAATATGATTAATAAATAATTTAAT 342

Qy 5567 TATTGTTTATAGCAAGACATATTTCTGAATGTTTTTAAATCTTTTAAATATATTTTATATATG 5626
Db TTTTATTTTATAAAGATTTAAATTTTAAATATTTGTAATTTATTTTATTTTATTTTATTTATAT 402

Qy 5627 TAGGTAATATTTTGTATCAATGCAATTTATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5680
Db ATCTATTTTATAAATAATTTATGTTGATTTATATATATATATATATATATATATATATATATATATAT 456

RESULT 12
US-08-998-416-288
Sequence 288, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter

APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
APPLICANT: Rebsch, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/398,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Weigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 288:
SEQUENCE CHARACTERISTICS:
LENGTH: 837 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1241RP
US-08-998-416-288

Query Match 0.8%; Score 55.2; DB 3; Length 837;
Best Local Similarity 49.2%; Pred. No. 0.0025;
Matches 174; Conservative 0; Mismatches 178; Indels 2; Gaps 1;
QY 5327 TCAGGCTCTAGACTTAAATCTTTGATTAATAAAGTGAATCTTATTTGTTATTAATGAGT 5386
DB 105 TTAATTAATTTGATTAATTAATACTTATTAATAATTAATTAATAATTAATTAATAATCA 164
QY 5387 GGAGACCTACCTCCATAATTAAGATAAACTCTTTTGGATTAATAATCAGAAATTTGCGCTT 5446
DB 165 TTAATTAATTTAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATG 224
QY 5447 TTTCTTCTCAAAATTAATACATAGTATGATATATATATATATATATATATATATATATAT 5506
DB 225 TCAAAATTTAAATAGTTATTAATAATTAATTAATAATTAATTAATTAATTAATTAATTA 284
QY 5507 TTAATGATATTAATAATAATGCGGGTCTTCAGGACTTTTCTCTATATATTAAGTA 5566
DB 285 TTAATAGATTAATCAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATA 342
QY 5567 TATGTTTTTATAGCAAGAACATATCTGAAGTTTTATAAATCTTTAAATATATATATG 5626
DB 343 TTTTATTAATTAAGAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 402
QY 5627 TAGGTAATATTTTGTATCAACAAGCATATTTTTTTCTCTCTCTCTCTCTCTCTCTCT 5680
DB 403 ATCTATTTTAAATAATATATGTTGATTTATATTAATTAATTAATTAATTAATTAATTA 456

RESULT 13

US-10-204-708-14
Sequence 14, Application US/10204708
Patent No. 6677731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
TITLE OF INVENTION: by Assessing DNA Methylation
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204,708
CURRENT FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 14
LENGTH: 6113
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-14

Query Match 0.8%; Score 53.8; DB 4; Length 6113;
Best Local Similarity 43.0%; Pred. No. 0.012;
Matches 473; Conservative 0; Mismatches 622; Indels 6; Gaps 4;
QY 5004 GGAATGATCTTTTCAAAATCATTAGCACAAATATTAAATATCTAAATAATTAAGAGATCCA 5063
DB 4889 GGAATGGTTGAGTTAATTTGTTAGGTATGTTAAATTTTTTAAGAAATTTGTTAAGTGG 4948
QY 5064 TACTTTCTGAGCTTTACAAATTAATTAAGTACTAAAGACAAAGAGATT---TCTTTTAA 5120
DB 4949 TGGTATCGTTTATTTTATATAGAGTGATAGATTTTGTAGTTGTTTATATATTTTA 5008
QY 5121 GAAATTTATAGCAATTTACTGCTTTATTTTAAATGCTTAAGCCAAAGTATCTGCACTTAGGTA 5180
DB 5009 TTAATATTTGGTATGTTAGTTTATTTTAAATTTTGTAGTTTGTAGTGGTGTGTTGGTA 5068
QY 5181 TACCTCTTTATGCCAATAATGATTTTAAATGAAGGCTCTTTTCAGATGA-ACCTTATGAA 5239
DB 5069 TTTTATTTTTTTTTTTTGTAGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5128
QY 5240 GGAATATCTGCTTTGTTATATGCGAGTTAGCAATCTGTTTCTAAAGTCTGTCAAAAT 5299
DB 5129 TAAATATAGTAAATTAATTTTAAATTTGTTGTTAGTTTATTAATAATTTGAATATATG 5188
QY 5300 GTATTTCTAGTGGCAAAACCAAGTTTGTAGGCTCTTAGACTTAAATTTCTTTGAATAAAA 5359
DB 5189 TGAATGATTAATTAATTAATAATTAAGATATAGATTTGTTGTTGTTGTTGTTGTTGTT 5248
QY 5360 CTGATAACTTATTTGTTAATTTGGAGTGGAGACCTACCTCCATATTAATAGATAAACTCTT 5419
DB 5249 TTTTATGTTATTTTTTTGTTAGTTAATAATTTTTTTTATTTTATTTTATTTAGTATATG 5308
QY 5420 TTGGATTAATAACAGAAATTTTCCCTTTTCTCTCAAAATTTATACATATGATGATTA 5479
DB 5309 AATTTGTTTTTCTTATTAATTTTTTTTTTATTTGAGATTTGTTTATATGATATATAT 5368
QY 5480 TATATCCACATATAGTTTTTCCCTGATTAAATGATTAATAATTTGCGGTGCTTC 5539
DB 5369 TATAATGTAATTTTTTGGATTTGGTTTTTTTATTTAGTATTAATGTTTGTAGATTTATA 5428

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2004, 02:10:43 ; Search time 1709.6 Seconds
(without alignments)
17604.315 Million cell updates/sec

Title: US-09-744-167-3

Perfect score: 6632

Sequence: 1 attccggccgggtagctc.....ttggaggcacattttgaagt 6632

Scoring table: IDENTITY NUC

Gapcp 10.0 , Gapext 1.0

Searched: 2947324 seqs, 2269024515 residues

Total number of hits satisfying chosen parameters: 5894648

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:**

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
- 3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq*
- 4: /cgn2_6/ptodata/1/pubpna/US05_PUBCOMB.seq*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
- 8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
- 17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq*
- 18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
- 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5290	79.8	5402	15	US-10-072-977-11
2	1514	22.8	1572	9	US-09-925-299-51
3	1514	22.8	1572	10	US-09-925-299-51
4	850	12.8	969	16	US-10-264-237-394
5	472.4	7.1	475	10	US-09-918-995-3724
6	444	6.7	455	10	US-09-930-213-158
7	389.4	5.9	391	10	US-09-918-995-7316
8	300.4	4.5	420	13	US-10-085-783A-29775
9	300.4	4.5	420	16	US-10-242-535A-29775
10	286	4.3	317	15	US-10-106-698-2133
11	199.6	3.0	393	9	US-09-960-352-13367
12	181.6	2.7	289	14	US-10-040-739-971
13	172	2.6	556	9	US-09-864-761-13497
14	168	2.5	168	9	US-09-864-761-30043
					Sequence 11, Appl
					Sequence 51, Appl
					Sequence 394, Appl
					Sequence 3724, Appl
					Sequence 158, Appl
					Sequence 7316, Appl
					Sequence 29775, A
					Sequence 2133, A
					Sequence 13367, A
					Sequence 971, Appl
					Sequence 13497, A
					Sequence 30043, A

c 15	124	1.9	387	9	US-09-880-107-5	Sequence 5, Appl
c 16	93.6	1.4	255	10	US-09-930-213-577	Sequence 577, Appl
c 17	88.8	1.3	3673778	15	US-10-312-841-1	Sequence 1, Appl
c 18	85.4	1.3	474	13	US-10-027-632-195942	Sequence 195942,
c 19	85.4	1.3	474	16	US-10-027-632-195942	Sequence 1721, Ap
c 20	82	1.2	11805	15	US-10-311-455-1721	Sequence 195942,
c 21	77.2	1.2	158001	17	US-10-211-179-11	GENERAL INFORMATI
c 22	75.6	1.1	7597	15	US-10-311-455-986	Sequence 986, App
c 23	75.2	1.1	7597	15	US-10-239-676-185	Sequence 185, App
c 24	75.2	1.1	7597	15	US-10-311-455-1995	Sequence 1995, Ap
c 25	75	1.1	11812	13	US-10-221-714A-468	Sequence 468, App
c 26	75	1.1	11812	13	US-10-239-676-210	Sequence 210, App
c 27	75	1.1	11812	15	US-10-311-455-2092	Sequence 2092, Ap
c 28	75	1.1	11812	15	US-10-240-453-306	Sequence 306, App
c 29	74	1.1	61020	13	US-10-221-714A-514	Sequence 514, App
c 30	73.2	1.1	5689	13	US-10-221-714A-148	Sequence 148, App
c 31	73.2	1.1	5689	15	US-10-239-676-90	Sequence 90, Appl
c 32	73.2	1.1	5689	15	US-10-240-453-100	Sequence 100, App
c 33	73	1.1	5920	15	US-10-311-455-490	Sequence 490, App
c 34	72.6	1.1	5527	15	US-10-311-455-290	Sequence 290, App
c 35	72.6	1.1	5527	15	US-10-240-452-38	Sequence 238, Ap
c 36	72.4	1.1	15548	15	US-10-311-455-2128	Sequence 2128, Ap
c 37	72.2	1.1	15399	15	US-10-311-455-1488	Sequence 1488, Ap
c 38	71.2	1.1	6681	15	US-10-311-455-128	Sequence 128, App
c 39	71.2	1.1	6681	15	US-10-240-452-4	Sequence 4, Appl
c 40	71.2	1.1	18218	15	US-10-311-455-1922	Sequence 1922, Ap
c 41	71	1.1	6112	15	US-10-311-455-462	Sequence 462, App
c 42	71	1.1	7498	15	US-10-311-455-230-	Sequence 230, App
c 43	70.8	1.1	6314	15	US-10-240-452-14	Sequence 14, Appl
c 44	70.6	1.1	17183	15	US-10-311-455-459	Sequence 459, App
c 45	70.2	1.1	15732	15	US-10-239-676-95	Sequence 95, Appl

ALIGNMENTS

RESULT 1

US-10-072-977-11
; Sequence 11, Application US/10072977
; Publication No. US20030152933A1
; GENERAL INFORMATION:
; APPLICANT: Barash et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PT057P1
; CURRENT APPLICATION NUMBER: US/10/072,977
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: PCT/US01/25288
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: US 60/225,215
; PRIOR FILING DATE: 2000-08-14
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 3.1
; SEQ ID NO 11
; LENGTH: 5402
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-072-977-11

Query Match	79.8%;	Score 5290;	DB 15;	Length 5402;
Best Local Similarity	99.5%;	Pred. No. 0;		
Matches 5317;	Conservative	0;	Mismatches 20;	Indels 5;
Gaps	1;			
QY	185	ATATCTCTCCAGGATGTTCTCAAGGCATACAGAAATTAATTCGTAATTAAGTCTGCAGG	244	
Db	48	AAAGCTCCGAGGGCTCTAGGAGGCATACAGAAATTAATTCGTAATTAAGTCTGCAGG	107	
QY	245	TAGGATGACAGTATTTTAAAGCAGCTGTGAGTCGACAACTCTTGATGATTT	304	
Db	108	TAGGATGACAGTATTTTAAAGCAGCTGTGAGTCGACAACTCTTGATGATTT	167	
QY	305	TGAACAGAACCCAGATGAACAGAGATTATCTCGAGATGTACAAATGCATATGATTTAA	364	
Db	168	TGAACAGAACCCAGATGAACAGAGATTATCTCGAGATGTACAAATGCATATGATTTAA	227	

QY	365	CCAC	TGCTCAG	TTTCTT	CAGAGT	TGGCTT	CTC	CAGCGG	AACTT	CTC	ATTG	CTCTCC	AAAA	AGA	424
DB	228	CCAC	TGCTCAG	TTTCTT	CAGAGT	TGGCTT	CTC	CAGCGG	AACTT	CTC	ATTG	CTCTCC	AAAA	AGA	287
QY	425	CAA	GAGT	CGGTAA	TATAGT	TGTGC	CTCAT	CAAA	CAAGCT	TATG	GAA	CAAA	TATG	GAA	484
DB	288	CAA	GAGT	CGGTAA	TATAGT	TGTGC	CTCAT	CAAA	CAAGCT	TATG	GAA	CAAA	TATG	GAA	347
QY	485	CCT	GAATG	AAAA	AAAC	CACTC	AAGG	CACTT	ACTT	CTTAT	CAAA	TAATG	AAAA	AAATG	544
DB	348	CCT	GAATG	AAAA	AAAC	CACTC	AAGG	CACTT	ACTT	CTTAT	CAAA	TAATG	AAAA	AAATG	407
QY	545	ACT	TGAT	CTTCTT	CTTCTG	TGGAT	TGGT	TACTT	CAGAT	GAAAT	CTC	AGCC	TGTAT	TATAT	604
DB	408	ACT	TGAT	CTTCTT	CTTCTG	TGGAT	TGGT	TACTT	CAGAT	GAAAT	CTC	AGCC	TGTAT	TATAT	467
QY	605	GGG	ACGAT	TGTAT	AAAC	CTAT	CTCT	GTGAT	CTCT	GAAT	AGT	GACAT	GGGT	AACTT	664
DB	468	GGG	ACGAT	TGTAT	AAAC	CTAT	CTCT	GTGAT	CTCT	GAAT	AGT	GACAT	GGGT	AACTT	527
QY	665	AAC	CAAT	AGT	GAAGA	GATAT	TAAAA	AAATTA	TGCC	AGAT	CAATTT	TAAGT	CTTAAT	GCAGA	724
DB	528	AAC	CAAT	AGT	GAAGA	GATAT	TAAAA	AAATTA	TGCC	AGAT	CAATTT	TAAGT	CTTAAT	GCAGA	587
QY	725	TTCC	TGAT	TGGAT	TGGAAT	TATCTT	CAGT	CTC	AGAT	ACTCC	CTGT	GTCTT	CTTCA	ACAGA	784
DB	588	TTCC	TGAT	TGGAT	TGGAAT	TATCTT	CAGT	CTC	AGAT	ACTCC	CTGT	GTCTT	CTTCA	ACAGA	647
QY	785	CCAT	GATAG	TGAT	ACTGT	CAGAG	AAC	ACAGAA	TGAT	ATCAG	TCTCGA	ATTTAC	AAATAG	844	
DB	648	CCAT	GATAG	TGAT	ACTGT	CAGAG	AAC	ACAGAA	TGAT	ATCAG	TCTCGA	ATTTAC	AAATAG	707	
QY	845	AGAA	ATCGG	AGGAAT	CAAA	GAATTTGG	GTAT	AAAA	AGTAG	ATAC	ACAC	CTTT	CAGAT	TCCCTA	904
DB	708	AGAA	ATCGG	AGGAAT	CAAA	GAATTTGG	GTAT	AAAA	AGTAG	ATAC	ACAC	CTTT	CAGAT	TCCCTA	767
QY	905	TAAT	TACAGT	GGAC	AGAAAT	TTTAA	AGAT	TTAA	AGAT	CTTT	TAAT	CAGT	TAGAT	CAAT	964
DB	768	TAAT	TACAGT	GGAC	AGAAAT	TTTAA	AGAT	TTAA	AGAT	CTTT	TAAT	CAGT	TAGAT	CAAT	827
QY	965	TGTT	GATTT	TAAC	ATGTC	ATCTG	CTTTG	ACT	CGA	CAAA	AGTTTCC	AAAA	TGTTT	CATG	1024
DB	828	TGTT	GATTT	TAAC	ATGTC	ATCTG	CTTTG	ACT	CGA	CAAA	AGTTTCC	AAAA	TGTTT	CATG	887
QY	1025	AGAC	AACT	TAC	AA	CAC	AGAG	CCAG	CCAT	GTGG	ATTACT	AA	AGAT	GTTCG	1084
DB	888	AGAC	AACT	TAC	AA	CAC	AGAG	CCAG	CCAT	GTGG	ATTACT	AA	AGAT	GTTCG	947
QY	1085	AGG	GAAT	GAT	GTGC	GAGT	CTATA	CTG	CGC	GAGA	TGTTT	AAAA	GAAG	GGCA	1144
DB	948	AGG	GAAT	GAT	GTGC	GAGT	CTATA	CTG	CGC	GAGA	TGTTT	AAAA	GAAG	GGCA	1007
QY	1145	AAG	TCTTT	GAC	CTG	CA	CCCTT	CG	AAAA	TATG	CTTAA	TGCTT	AAAT	GAT	1204
DB	1008	AAG	TCTTT	GAC	CTG	CA	CCCTT	CG	AAAA	TATG	CTTAA	TGCTT	AAAT	GAT	1067
QY	1205	AAG	AGAT	GAAAA	TTTT	CAAA	TACT	CTG	ACT	TTTT	CTCTT	CAG	GAAG	ATG	1264
DB	1068	AAG	AGAT	GAAAA	TTTT	CAAA	TACT	CTG	ACT	TTTT	CTCTT	CAG	GAAG	ATG	1127
QY	1265	ACA	ATCTG	CA	CA	GA	AG	ACT	CA	AAAA	AGTTT	AG	AGAT	TAAT	1324
DB	1128	ACA	ATCTG	CA	CA	GA	AG	ACT	CA	AAAA	AGTTT	AG	AGAT	TAAT	1187
QY	1325	TTCC	TCTT	CAG	CTTT	TAC	ATG	TTTT	CC	AGTA	AA	GAT	TG	CCCT	1384
DB	1188	TTCC	TCTT	CAG	CTTT	TAC	ATG	TTTT	CC	AGTA	AA	GAT	TG	CCCT	1247
QY	1385	TGCG	T	CGGT	CT	AT	GTG	TG	GCAT	TT					

QY	1445	TCAGCATGAA	CATAAAGAN	TAATATAT	ACAAGATG	CGAGTGA	CTATACAT	GAAGAA	TATACAGAA	1501		
DB	1308	TCAGCATGAA	CATAAAGAN	TAATATAT	ACAAGATG	CGAGTGA	CTATACAT	GAAGAA	TATACAGAA	1367		
QY	1505	CAGTGTCTT	CTAGTG	GGGAA	CCATTCA	AAAGAGA	ATGATCTTT	TGAAAC	CAGGAAA	ATG 1564		
DB	1368	CAGTGTCTT	CTAGTG	GGGAA	CCATTCA	AAAGAGA	ATGATCTTT	TGAAAC	CAGGAAA	ATG 1427		
QY	1565	TAAAGCAT	ACTCCTT	CAGTCA	TTAATTG	GAAGG	GATGGA	ACAGAA	GATAGAT	CCTGA 1624		
DB	1428	TAAAGCAT	ACTCCTT	CAGTCA	TTAATTG	GAAGG	GATGGA	ACAGAA	GATAGAT	CCTGA 1487		
QY	1625	CCAGACAG	TAAATCAG	ACTGAGT	CTTTGG	ATGTTGG	TGACAC	CAGTCTC	ACAGTTGT	TAGA 1684		
DB	1488	CCAGACAG	TAAATCAG	ACTGAGT	CTTTGG	ATGTTGG	TGACAC	CAGTCTC	ACAGTTGT	TAGA 1547		
QY	1685	ATCTCAAG	AGGGGCTT	CTGGCA	CTCATGT	CCACAG	AGTCTTCT	GATTTGT	GTCAAG	GGTTT 1744		
DB	1548	ATCTCAAG	AGGGGCTT	CTGGCA	CTCATGT	CCACAG	AGTCTTCT	GATTTGT	GTCAAG	GGTTT 1607		
QY	1745	TATTAAAT	ACTTTTT	CAAGCA	ATGATG	GATGG	CAAGACTT	AGATTA	CTTTTAAT	TATGA 1804		
DB	1608	TATTAAAT	ACTTTTT	CAAGCA	ATGATG	GATGG	CAAGACTT	AGATTA	CTTTTAAT	TATGA 1667		
QY	1805	TGAAGG	CGCAAAA	AGTGG	CCACCTA	ATTAGT	GATGCTG	GAAC	TTGATG	CCCTTCTGAC	AGA 1864	
DB	1668	TGAAGG	CGCAAAA	AGTGG	CCACCTA	ATTAGT	GATGCTG	GAAC	TTGATG	CCCTTCTGAC	AGA 1727	
QY	1865	ACAGTAT	CTTCAG	ACCCTA	CACTAA	ATAAGT	CTTTTGA	AGAAA	ATGTAAT	GACTCTAA	ATC 1924	
DB	1728	ACAGTAT	CTTCAG	ACCCTA	CACTAA	ATAAGT	CTTTTGA	AGAAA	ATGTAAT	GACTCTAA	ATC 1787	
QY	1925	GCAATG	CAATCAG	ATAGAT	GAAAG	CGCTPAG	ATGATG	GGAAC	ATCAAT	TAATATAT	ATTT 1984	
DB	1788	GCAATG	CAATCAG	ATAGAT	GAAAG	CGCTPAG	ATGATG	GGAAC	ATCAAT	TAATATAT	ATTT 1847	
QY	1985	CAATG	CAAGAC	GAGGCTA	TTGGG	AAAGTCTA	GTGTTA	TATATAT	TAATTTGT	GAAAC	AGT 2044	
DB	1848	CAATG	CAAGAC	GAGGCTA	TTGGG	AAAGTCTA	GTGTTA	TATATAT	TAATTTGT	GAAAC	AGT 1907	
QY	2045	TGATAA	CAAAAT	ACAATAG	AAAA	TGGCCTTCTT	TAGG	AGAAA	AAAGCA	CACTATT	CCAGT 2104	
DB	1908	TGATAA	CAAAAT	ACAATAG	AAAA	TGGCCTTCTT	TAGG	AGAAA	AAAGCA	CACTATT	CCAGT 1967	
QY	2105	TCAAC	AGGGT	TACCTAC	CAAGTAA	GCTGAG	ATTA	CAAAAT	CAATAT	CAGTCTCT	GATAT 2164	
DB	1968	TCAAC	AGGGT	TACCTAC	CAAGTAA	GCTGAG	ATTA	CAAAAT	CAATAT	CAGTCTCT	GATAT 2027	
QY	2165	TAACAGT	CAATCTG	TTGG	AGGGCC	AGACCTA	AGCAA	TTGTTTAG	CGCTTCC	ATCAAG	AAAC 2224	
DB	2028	TAACAGT	CAATCTG	TTGG	AGGGCC	AGACCTA	AGCAA	TTGTTTAG	CGCTTCC	ATCAAG	AAAC 2087	
QY	2225	AAGGAGT	TCAAAG	CACTG	AAATGA	AGCCAG	ATGTTCC	CAGATAC	AAATAG	AAAGTGA	ACCCAG 2284	
DB	2088	AAGGAGT	TCAAAG	CACTG	AAATGA	AGCCAG	ATGTTCC	CAGATAC	AAATAG	AAAGTGA	ACCCAG 2147	
QY	2285	CACAGCAG	ATACCG	TGTTTCC	AAATCA	CTTGTG	CTATAG	ATTTAC	AGCTGAT	CCACAG	GT 2344	
DB	2148	CACAGCAG	ATACCG	TGTTTCC	AAATCA	CTTGTG	CTATAG	ATTTAC	AGCTGAT	CCACAG	GT 2207	
QY	2345	TAGCTT	CAACT	TAATAT	CATATG	ATAAG	ATTAATCT	GAGGTG	GAATCTAG	TTTTCGT	2404	
DB	2208	TAGCTT	CAACT	TAATAT	CATATG	ATAAG	ATTAATCT	GAGGTG	GAATCTAG	TTTTCGT	2267	
QY	2405	AAC	TGCAAA	TGAAG	ATTC	TGTAC	CTGAAA	CACTTG	CAAAAG	AGCGT	TGGTTTGG	GCCA 2464
DB	2268	AAC	TGCAAA	TGAAG	ATTC	TGTAC	CTGAAA	CACTTG	CAAAAG	AGCGT	TGGTTTGG	GCCA 2327
QY	2465	GAAC	AGCCTACT	TGGT	CTTCTG	ATTCAG	AGCTCC	AACTGT	ATAG	ACTGCC	AGTCAA 2524	
DB	2328	GAAC	AGCCTACT	TGGT	CTTCTG	ATTCAG	AGCTCC	AACTGT	ATAG	ACTGCC	AGTCAA 2387	
QY	2525	ATT	TACTTTT	TACAA	ACGG	CGAC	ACCAAT	TGCG	AGCATG	CGGAA	AGTATTTTGT	GGTGT 2584


```
QY 4745 GCCTCAGCATTAATGATCTTGATAGTCTCTGATACCTGTGATCCATGCTGGAC 4804
DB 4608 GCCTCAGCATTAATGATCTTGATAGTCTCTGATACCTGTGATCCATGCTGGAC 4667
QY 4805 CTCCTCAACTCTAGTTTACCAATAGAAATAGATAGTCTGTTTTCATATAGAACATCTTTT 4864
DB 4668 CTCCTCAACTCTAGTTTACCAATAGAAATAGATAGTCTGTTTTCATATAGAACATCTTTT 4727
QY 4865 TTAGTGAAGAATGCGCATATATAGATATGCAACCTAAATTTGTTTAAACTAACCTCCAGC 4924
DB 4728 TTAGTGAAGAATGCGCATATATAGATATGCAACCTAAATTTGTTTAAACTAACCTCCAGC 4787
QY 4925 ACTAAGCTGAAATGCCACAAACACCTAAAGTATATAATATGCTGATTTTGAACACAT 4984
DB 4788 ACTAAGCTGAAATGCCACAAACACCTAAAGTATATAATATGCTGATTTTGAACACAT 4847
QY 4985 AAGCTTTGCTCTTTAGGAGGAATGATCTTTTCAAAATCATTAGCACAAATATTTAAATATC 5044
DB 4848 AAGCTTTGCTCTTTAGGAGGAATGATCTTTTCAAAATCATTAGCACAAATATTTAAATATC 4907
QY 5045 TAAATAATTAAGATGCTACTTTCTGATCTTTTACAAATCATTAGCACAAATATTTAAATATC 5104
DB 4908 TAAATAATTAAGATGCTACTTTCTGATCTTTTACAAATCATTAGCACAAATATTTAAATATC 4962
QY 5105 CAAGGATTTCTTTTAAAGAAATTTATAGCATTTACTGTTTATTTAAATGCTTAAGCCAAAG 5164
DB 4963 CAAGGATTTCTTTTAAAGAAATTTATAGCATTTACTGTTTATTTAAATGCTTAAGCCAAAG 5022
QY 5165 TATCTGCACTTAGTATACCTCTTTATGCCAAATATGATTTTAAATGAGGCTCTTTTTCAG 5224
DB 5023 TATCTGCACTTAGTATACCTCTTTATGCCAAATATGATTTTAAATGAGGCTCTTTTTCAG 5082
QY 5225 ATGTAACCTTTATGAAGAAATATCTGCTTGTATATGCCAGTTAGAAATCTGTTTCT 5284
DB 5083 ATGTAACCTTTATGAAGAAATATCTGCTTGTATATGCCAGTTAGAAATCTGTTTCT 5142
QY 5285 AAAGTCTGCAAAATTTATTCAGTGGCAAAAACCAAGTTTTCCTCTCAAAATTTATA 5344
DB 5143 AAAGTCTGCAAAATTTATTCAGTGGCAAAAACCAAGTTTTCCTCTCAAAATTTATA 5202
QY 5345 ATTCCTTCAATAAATCTGATTAATCTTATTTGATATTTGGAGTGGAGACCTACCTCCATAA 5404
DB 5203 ATTCCTTCAATAAATCTGATTAATCTTATTTGATATTTGGAGTGGAGACCTACCTCCATAA 5262
QY 5405 TTAGATAAATCTTTTGGATTAATCAGAAATTTTGCCTTTTCTCTCAAAATTTATA 5464
DB 5263 TTAGATAAATCTTTTGGATTAATCAGAAATTTTGCCTTTTCTCTCAAAATTTATA 5322
QY 5465 CATATGATGATTAATATATATCCACATATATAGTTTTCCTGATTAATGATATTAAT 5524
DB 5323 CATATGATGATTAATATATATCCACATATATAGTTTTCCTGATTAATGATATTAAT 5382
QY 5525 AA 5526
DB 5383 AA 5384
```

```
RESULT 2
US-09-925-299-51
; Sequence 51, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1555)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-51

Query Match 22.8%; Score 1514; DB 9; Length 1572;
Best Local Similarity 99.3%; Pred. No. 6.2e-304; Mismatches 3; Indels 1; Gaps 1;
Matches 1522; Conservative 7;

QY 4865 TTAGTGAAGAATGCGCATATATACATATTCGCAACCTAAATTTGTTTAAACTAACCTCCAGC 4924
DB 1 TNAGTGAAGAATGCGCATATATACATATTCGCAACCTAAATTTGTTTAAACTAACCTCCAGC 60
QY 4925 ACTAAGCTGAAATGCCACAAACACCTAAAGTATATAATATGCTGATTTTGAACACAT 4984
DB 61 ACTAAGCTGAAATGCCACAAACACCTAAAGTATATAATATGCTGATTTTGAACACAT 120
QY 4985 AAGCTTTGCTCTTTAGGAGGAATGATCTTTTCAAAATCATTAGCACAAATATTTAAATATC 5044
DB 121 AAGCTTTGCTCTTTAGGAGGAATGATCTTTTCAAAATCATTAGCACAAATATTTAAATATC 180
QY 5045 TAAATAATTAAGATGCTACTTTCTGATCTTTTCAAAATCATTAGCACAAATATTTAAATATC 5104
DB 181 TAAATAATTAAGATGCTACTTTCTGATCTTTTCAAAATCATTAGCACAAATATTTAAATATC 240
QY 5105 CAAGGATTTCTTTTAAAGAAATTTATAGCATTTACTGTTTATTTAAATGCTTAAGCCAAAG 5164
DB 241 CAAGGATTTCTTTTAAAGAAATTTATAGCATTTACTGTTTATTTAAATGCTTAAGCCAAAG 300
QY 5165 TATCTGCACTTAGTATACCTCTTTATGCCAAATATGATTTTAAATGAGGCTCTTTTTCAG 5224
DB 301 TATCTGCACTTAGTATACCTCTTTATGCCAAATATGATTTTAAATGAGGCTCTTTTTCAG 360
QY 5225 ATGTAACCTTTATGAAGAAATATCTGCTTGTATATGCCAGTTAGAAATCTGTTTCT 5284
DB 361 ATGTAACCTTTATGAAGAAATATCTGCTTGTATATGCCAGTTAGAAATCTGTTTCT 420
QY 5285 AAAGTCTGCAAAATTTATTCAGTGGCAAAAACCAAGTTTTCCTCTCAAAATTTATA 5344
DB 421 AAAGTCTGCAAAATTTATTCAGTGGCAAAAACCAAGTTTTCCTCTCAAAATTTATA 480
QY 5345 ATTCCTTCAATAAATCTGATTAATCTTATTTGATATTTGGAGTGGAGACCTACCTCCATAA 5404
DB 481 ATTCCTTCAATAAATCTGATTAATCTTATTTGATATTTGGAGTGGAGACCTACCTCCATAA 540
QY 5405 TTAGATAAATCTTTTGGATTAATCAGAAATTTTGCCTTTTCTCTCAAAATTTATA 5464
DB 541 TTAGATAAATCTTTTGGATTAATCAGAAATTTTGCCTTTTCTCTCAAAATTTATA 600
QY 5465 CATATGATGATTAATATATATCCACATATATAGTTTTCCTGATTAATGATATTAAT 5524
DB 601 CATATGATGATTAATATATATCCACATATATAGTTTTCCTGATTAATGATATTAAT 660
QY 5525 AATTCGGGTGCTTCAGGACTTTTTCCTCTATATTTAAAGTATATTTGTTTATAGCAAG 5584
DB 661 AATTCGGGTGCTTCAGGACTTTTTCCTCTATATTTAAAGTATATTTGTTTATAGCAAG 720
QY 5585 AACATATTTCTGAATGTTTATATAATCTTTTAAATATTTATGATAGTAAATATTTTGTAT 5644
DB 721 AACATATTTCTGAATGTTTATATAATCTTTTAAATATTTATGATAGTAAATATTTTGTAT 780
QY 5645 CACAATGATATTTTTCCTCTTCCCTTCCAACTATACCACTGATTTTACCCTT 5704
DB 781 CACAATGATATTTTTCCTCTTCCCTTCCAACTATACCACTGATTTTACCCTT 839
```

```
QY 5705 CTAAGAGTGACTGACGACGGCCAGATGACCCCTGAAGTAGTCATTATGTAGCAATTAAT 5764
DB 840 CTAAGAGTGACTGACGACGGCCAGATGACCCCTGAAGTAGTCATTATGTAGCAATTAAT 899
QY 5765 GAAGCCTGAAACACAGGTTTTTACTTTCACCTTTAAGCTTGAAGAAATTTCTTGGCAACTTC 5824
DB 900 GAAGCCTGAAACACAGGTTTTTACTTTCACCTTTAAGCTTGAAGAAATTTCTTGGCAACTTC 959
QY 5825 GCATATTTCTTATGACACGTGTTATAGTATTAATTTAAATGAACATAATTTACTTTTGGCA 5884
DB 960 GCATATTTCTTATGACACGTGTTATAGTATTAATTTAAATGAACATAATTTACTTTTGGCA 1019
QY 5885 TATTTTAAATTTCTTATGAGTATTTATTTTATACAGGATATTAACATAAGCTTAAAT 5944
DB 1020 TATTTTAAATTTCTTATGAGTATTTATTTTATACAGGATATTTAAATGAACATAATTTACTTTTGGCA 1079
QY 5945 CCTATGATTTTGAATTTCTTACAGAGCTTTCTCTTTACTTTCAACACGCAAAAAAGTGGG 6004
DB 1080 CCTATGATTTTGAATTTCTTACAGAGCTTTCTCTTTACTTTCAACACGCAAAAAAGTGGG 1139
QY 6005 GGGCATATTTAGTCTCTGCTCAATTTAAGTTATGTAAATAATTTAATCATTTATTTGATGCT 6064
DB 1140 GGGCATATTTAGTCTCTGCTCAATTTAAGTTATGTAAATAATTTAATCATTTATTTGATGCT 1199
QY 6065 TTAAACATTTCTCATGTATATATATTTTGTATCAAAAAACATCATATATTTCAAGAA 6124
DB 1200 TTAAACATTTCTCATGTATATATATTTTGTATCAAAAAACATCATATATTTCAAGAA 1259
QY 6125 AAAGAAATTTGTTAAATAGCCCTGTTTAAAGAAATTTTATGAAGCATCTCAACTTG 6184
DB 1260 AAAGAAATTTGTTAAATAGCCCTGTTTAAAGAAATTTTATGAAGCATCTCAACTTG 1319
QY 6185 AAGATCAAGTCAAGTATTAATCTCAGGATCTGAGGTCTCAAGCTAGGAGACCTGAGAAAT 6244
DB 1320 AAGATCAAGTCAAGTATTAATCTCAGGATCTGAGGTCTCAAGCTAGGAGACCTGAGAAAT 1379
QY 6245 TTAAATCAGTTTGGCATATATAGTTTGGACTGAATCAATCTGTAGTACTTAGCCAAAGAC 6304
DB 1380 TTAAATCAGTTTGGCATATATAGTTTGGACTGAATCAATCTGTAGTACTTAGCCAAAGAC 1439
QY 6305 AATTTGGAGGAGAAATATCAGCCTTCTGGAAGTAGTACTTCTGAAACAATGTAAAGTCTC 6364
DB 1440 AATTTGGAGGAGAAATATCAGCCTTCTGGAAGTAGTACTTCTGAAACAATGTAAAGTCTC 1499
QY 6365 GCAGATATTTCAATAAAATGGCAACCTGTTTAA 6397
DB 1500 GCAGATATTTCAATAAAATGGCAACCTGTTTAA 1532

RESULT 3
US-09-925-299-51
; Sequence 51, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PAL02
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2)
```

```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1555)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-51
```

```
Query Match 22.8%; Score 1514; DB 10; Length 1572;
Best Local Similarity 99.3%; Pred. No. 6.2e-304;
Matches 1522; Conservative 7; Mismatches 3; Indels 1; Gaps 1;
```

```
QY 4865 TTAGTGAAGAATGTGCCATATTTACATATTGCAACCTAATTTGTTAAAACTAACTCCAGC 4924
DB 1 TNAGTGAAGAATGTGCCATATTTACATATTGCAACCTAATTTGTTAAAACTAACTCCAGC 60
QY 4925 ACTAAAGCTTGAATGTGCCACAAAACACATAAAGTATAATATGCTGATTTTGAACACAT 4984
DB 61 ACTAAAGCTTGAATGTGCCACAAAACACATAAAGTATAATATGCTGATTTTGAACACAT 120
QY 4985 AAGCTTTGCTTTTAGGAGGAATGATCTTTCAAAATCATAGCACAATATTTAAATATC 5044
DB 121 AAGCTTTGCTTTTAGGAGGAATGATCTTTCAAAATCATAGCACAATATTTAAATATC 180
QY 5045 TAAAAATTTAAGAGATCCATACCTTTCTGTAGCTTTACAATTTAATTAAGTACTAAAAAGA 5104
DB 181 TAAAAATTTAAGAGATCCATACCTTTCTGTAGCTTTACAATTTAATTAAGTACTAAAAAGA 240
QY 5105 CAAGGATTTCTTTTAAAGAAATTTATAGCAATTTACTGTGTTATTTAAATGCTAAGCCAAAG 5164
DB 241 CAAGGATTTCTTTTAAAGAAATTTATAGCAATTTACTGTGTTATTTAAATGCTAAGCCAAAG 300
QY 5165 TATCTGCACCTTAGGTATACCTTTTATGCCAATAATGATTTTAAATGAAGCTCTTTTCAG 5224
DB 301 TATCTGCACCTTAGGTATACCTTTTATGCCAATAATGATTTTAAATGAAGCTCTTTTCAG 360
QY 5225 ATGTAACTTTATGAAGAAATATCTGCTTTGTGTATATGCCAGTTAGAATCTGTTTCT 5284
DB 361 ATGTAACTTTATGAAGAAATATCTGCTTTGTGTATATGCCAGTTAGAATCTGTTTCT 420
QY 5285 AAAGTCTGCAAAATGATTTTCAAGTGCACAAAACACAGTTTTCAGTCTTAGCTTATA 5344
DB 421 AAAGTCTGCAAAATGATTTTCAAGTGCACAAAACACAGTTTTCAGTCTTAGCTTATA 480
QY 5345 ATTCTTTGAATAAACTGATAAATTTTGTATAAATTTGGAGTGGAGACCTTCCATAA 5404
DB 481 ATTCTTTGAATAAACTGATAAATTTTGTATAAATTTGGAGTGGAGACCTTCCATAA 540
QY 5405 TTAGATAAACTCTTTTGGATATATAATCAGAAATTTTGCCTTTTCTCTCAAAATTTATA 5464
DB 541 TTAGATAAACTCTTTTGGATATATAATCAGAAATTTTGCCTTTTCTCTCAAAATTTATA 600
QY 5465 CATATGTATGTATATATATATCCATATATAGTTTTCCTCTGATTAAATGGATATTAAT 5524
DB 601 CATATGTATGTATATATATATCCATATATAGTTTTCCTCTGATTAAATGGATATTAAT 660
QY 5525 AATTGCGGGTCTTCAGGACTTTTGTCTATATTTTAAAGTATATTTTATAGCAAG 5584
DB 661 AATTGCGGGTCTTCAGGACTTTTGTCTATATTTTAAAGTATATTTTATAGCAAG 720
QY 5585 AACATATCTGAATGTTTTTATAATCTTTAATTTATATGTAGTAAATATTTTCTAT 5644
DB 721 AACATATCTGAATGTTTTTATAATCTTTAATTTATATGTAGTAAATATTTTCTAT 780
QY 5645 CACAATGCAATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5704
DB 781 CACAATGCAATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 839
QY 5705 CTAAGAGTGACTGACGACGGCCAGATGACCCCTGAAGTAGTCATTATGTAGCAATTAAT 5764
DB 840 CTAAGAGTGACTGACGACGGCCAGATGACCCCTGAAGTAGTCATTATGTAGCAATTAAT 899
QY 5765 GAAGCCTGAAACACAGGTTTTTACTTTCACCTTTAATCTTTAGAAATTTCTTGGCAACTTC 5824
DB 900 GAAGCCTGAAACACAGGTTTTTACTTTCACCTTTAATCTTTAGAAATTTCTTGGCAACTTC 959
```


Db 897 NTNTGAAATGGAAAAACCGCTTAAATTTCCCGGAAAA 938

RESULT 5

US-09-918-995-3724
; Sequence 3724, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3724
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-3724

Query Match 7.1%; Score 472.4; DB 10; Length 475;
Best Local Similarity 99.8%; Pred. No. 8e-88;
Matches 473; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	3496	ACTTTTCAACCAATGGATTCGATGCTTGGGACGAGCAGAAATATTATTCTATTGTTAT	3555
Db	1	ACTTTTCAACCAATGGATTCGATGCTAGGACGAGCAGAAATATTATTCTATTGTTAT	60
Qy	3556	GTTCGCCAAATGAAGATACCTATCTCAAGGACATCTTCAGACATATTATACCCATATA	3615
Db	61	GTTCGCCAAATGAAGATACCTATCTCAAGGACATCTTCAGACATATTATACCCATATA	120
Qy	3616	AGGATGCTCTAAAGGAAATACATAGAAACCTTCGACAAATATTACCTTTACTGAGATT	3675
Db	121	AGGATGCTCTAAAGGAAATACATAGAAACCTTCGACAAATATTACCTTTACTGAGATT	180
Qy	3676	TTCTCAGTAGCAAGGATCAGGAGGATTCCTGTTTATTACCTACTTTTCAGAACTTG	3735
Db	181	TTCTCAGTAGCAAGGATCAGGAGGATTCCTGTTTATTACCTACTTTTCAGAACTTG	240
Qy	3736	ATGATCTCTCATACAGATATCTTTCTTTGTTGGAATCTTATCCAGAGCTTGAGA	3795
Db	241	ATGATCTCTCATACAGATATCTTTCTTTGTTGGAATCTTATCCAGAGCTTGAGA	300
Qy	3796	TTCCCTGGGCAAGGTTTTCCTATGCGTTTAAATGTTGAGATTGGGTGCAGAAATATAAG	3855
Db	301	TTCCCTGGGCAAGGTTTTCCTATGCGTTTAAATGTTGAGATTGGGTGCAGAAATATAAG	360
Qy	3856	CATATCTCTCTCTCTAACAAGCATCAGAGCGCGAAACCTCTTTTTCGGAGAAATAGGAC	3915
Db	361	CATATCTCTCTCTCTAACAAGCATCAGAGCGCGAAACCTCTTTTTCGGAGAAATAGGAC	420
Qy	3916	ACACTATTATGAACTTACTTGTGTTGACCTTCGAAATACCAGTATACCTTTGCATA	3969
Db	421	ACACTATTATGAACTTACTTGTGTTGACCTTCGAAATACCAGTATACCTTTGCATA	474

RESULT 6

US-09-930-213-158/c
; Sequence 158, Application US/09930213
; Publication No. US20030170625A1
; GENERAL INFORMATION:
; APPLICANT: ROSENTHAL, ANDRE
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHAFER, REINHARD
; APPLICANT: ZUBER, JOHANNES
; APPLICANT: TCHE-NITSE, OLEG
; APPLICANT: GRIPS, MARTIN
; APPLICANT: HELLMER, MARTIN

; APPLICANT: SCHMITZ, ANNE-CHANTAL
; APPLICANT: SEMS, CHRISTINE
; TITLE OF INVENTION: DETECTION OF DIFFERENTIAL GENE EXPRESSIONS
; FILE REFERENCE: ALBRE-14
; CURRENT APPLICATION NUMBER: US/09/930,213
; CURRENT FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: DE 10004102.7
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 885
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 158
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-930-213-158

Query Match 6.7%; Score 444; DB 10; Length 455;
Best Local Similarity 99.8%; Pred. No. 6.2e-82;
Matches 455; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy	5337	GACTTATAATCTTTGAATAAACTGATAACTTATTGTTGTTATTAATTTGGAGTGGAGACTAC	5396
Db	455	GACTTATAATCTTTGAATAAACTGATAACTTATTGTTGTTATTAATTTGGAGTGGAGACTAC	396
Qy	5397	CTCCATAATTAGATAAACTCTTTTGGATTATAATCAGAAATTTTGCTTTTCTTCTCA	5456
Db	395	CTCCATAATTAGATAAACTCTTTTGGATTATAATCAGAAATTTTGCTTTTCTTCTCA	336
Qy	5457	AATTATTACATATGTTGTTATTAATATCCACATATATAGTTTCCCTGATTAAATGGAT	5516
Db	335	AATTATTACATATGTTGTTATTAATATCCACATATATAGTTTCCCTGATTAAATGGAT	276
Qy	5517	ATTAAATAATTCGGGTGCTTCAGACATTTTGGCTTCTATATTAAAGTATATTGTTTT	5576
Db	275	ATTAAATAATTCGGGTGCTTCAGACATTTTGGCTTCTATATTAAAGTATATTGTTTT	216
Qy	5577	ATAGCAAGACATATCTGAATCTTTTATAAACTTTATAAATTTTATATAGTAAATAT	5636
Db	215	ATAGCAAGACATATCTGAATCTTTTATAAATTTTATAAATTTTATATAGTAAATAT	156
Qy	5637	TTTGTATCAATATGCAATATTTTTCCTCTCTCTTCCCTTCCAACTATACCACTGTAAT	5696
Db	155	TTTGTATCAATATGCAATATTTTTCCTCTCTCTTCCCTTCCAACTATACCACTGTAAT	97
Qy	5697	TACCACCTCTAAGAGTCACTGACGCGGCGACATGACCTTGAAGTAGTCATTATGTAG	5756
Db	96	TACCACCTCTAAGAGTCACTGACGCGGCGACATGACCTTGAAGTAGTCATTATGTAG	37
Qy	5757	CAATAAATGAAGCTGAAACAGGTTTTTTTACTTCC 5792	
Db	36	CAATAAATGAAGCTGAAACAGGTTTTTTTACTTCC 1	

RESULT 7

US-09-918-995-7316
; Sequence 7316, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7316
; LENGTH: 391
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-7316

Query Match 5.9%; Score 389.4; DB 10; Length 391;
Best Local Similarity 99.7%; Pred. No. 1.3e-70;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4639 CTCACTGAAACTCTAAAGAGTAATGGATGAATAAATGGACTCAGAGTTTCCATTG 4698
DB 1 CTCACTGAAACTCTAAAGAGTAATGGATGAATAAATGGACTCAGAGTTTCCATTG 60

QY 4699 AACTGATATGGTGAATTTCAAGGAGGATCTGAAGGCCAACTTCTGCTCAGATTATC 4759
DB 61 AACTGATATGGTGAATTTCAAGGAGGATCTGAAGGCCAACTTCTGCTCAGATTATC 120

QY 4759 TAAATGATCTGATAGTCTCTGATACCTGTGATCCATGGTGGACCTCCCACTCTAGTT 4818
DB 121 TAAATGATCTGATAGTCTCTGATACCTGTGATCCATGGTGGACCTCCCACTCTAGTT 180

QY 4819 TACCATTAGAAATAGAAATAGTGTGTTTTCATTATAGAACATCTTTTATAGTGAAGAATG 4878
DB 181 TACCATTAGAAATAGAAATAGTGTGTTTTCATTATAGAACATCTTTTATAGTGAAGAATG 240

QY 4879 TGCATATATACATATGCAACCTAATTTGTTAACTAACTCCAGCACTAAAGCTGAAT 4938
DB 241 TGCATATATACATATGCAACCTAATTTGTTAACTAACTCCAGCACTAAAGCTGAAT 300

QY 4939 GCCACAAACACTAAAGAGTATAATATGCTGATTTTGAACACATAGCTTTGCTCTTT 4998
DB 301 GCCACAAACACTAAAGAGTATAATATGCTGATTTTGAACACATAGCTTTGCTCTTT 360

QY 4999 AGCAGGAATGATCTTTTCAATCATTAGCA 5029
DB 361 AGCAGGAATGATCTTTTCAATCATTAGCA 391

RESULT 8

US-10-085-783A-29775
; Sequence 29775, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29775
; LENGTH: 420
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (419)..(419)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-29775

Query Match 4.5%; Score 300.4; DB 13; Length 420;
Best Local Similarity 94.7%; Pred. No. 4.1e-52;
Matches 321; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

QY 185 ATATCTCTCCAGGATGTTCTCAAGGCATACAAAGATTAATAATCTGAATAAGTCTCAGG 244
DB 82 AAGCTCCGAGGGGCTGTAGGGGCAATACAGAAATTAATTTCTGAATAAGTCTCAGG 141

QY 245 TAGGATGGACAGTTATTTAAAGCAGCTGTCAAGTGTGAGCAAACTCCTTGATGATTT 304

DB 142 TAGGATGGACAGTTATTTAAAGCAGCTGTCAAGTGTGAGCAAACTCCTTGATGATTT 201

QY 305 TGAACAGAACCCAGATGAACAGATTATCTCGCAGATGTACAAAATGCATATGATTTCTAA 364

DB 202 TGAACAGAACCCAGATGAACAGATTATCTCGCAGATGTACAAAATGCATATGATTTCTAA 261

QY 365 CCAGTGTCTCAGTTTCTTTCAGAGTTGGCTTCTCCACAGCGAACTTTCATTCGCCAAAAGA 424

DB 262 CCAGTGTCTCAGTTTCTTTCAGAGTTGGCTTCTCCACAGCGAACTTTCATTCGCCAAAAGA 321

QY 425 CCAAGAGTGGTTAATAA-GTTGTGCTCATCAAGAAACAGCTATGGAACAAATGAGAGTT 483

DB 322 CCAAGAGTGGTTAATAA-GTTGTGCTCATCAAGAAACAGCTATGGAACAAATGAGAGTT 381

QY 484 CCCTGAATGAAAAACACTCAAGGAGCTTACTTCTATAC 522

DB 382 CCCTGAATGAAAAACACTCAAGGAGCTTACTTCTATNC 420

RESULT 9

US-10-242-535A-29775
; Sequence 29775, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29775
; LENGTH: 420
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (419)..(419)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-29775

Query Match 4.5%; Score 300.4; DB 16; Length 420;
Best Local Similarity 94.7%; Pred. No. 4.1e-52;
Matches 321; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

QY 185 ATATCTCTCCAGGATGTTCTCAAGGCATACAAAGATTAATAATCTGAATAAGTCTCAGG 244
DB 82 AAGCTCCGAGGGGCTGTAGGGGCAATACAGAAATTAATTTCTGAATAAGTCTCAGG 141

QY 245 TAGGATGGACAGTTATTTAAAGCAGCTGTCAAGTGTGAGCAAACTCCTTGATGATTT 304

DB 142 TAGGATGGACAGTTATTTAAAGCAGCTGTCAAGTGTGAGCAAACTCCTTGATGATTT 201

QY 305 TGAACAGAACCCAGATGAACAGATTATCTCGCAGATGTACAAAATGCATATGATTTCTAA 364

DB 202 TGAACAGAACCCAGATGAACAGATTATCTCGCAGATGTACAAAATGCATATGATTTCTAA 261

QY 365 CCAGTGTCTCAGTTTCTTTCAGAGTTGGCTTCTCCACAGCGAACTTTCATTCGCCAAAAGA 424

DB 262 CCAGTGTCTCAGTTTCTTTCAGAGTTGGCTTCTCCACAGCGAACTTTCATTCGCCAAAAGA 321

QY 425 CCAAGAGTGGTTAATAA-GTTGTGCTCATCAAGAAACAGCTATGGAACAAATGAGAGTT 483

DB 322 CCAAGAGTGGTTAATAA-GTTGTGCTCATCAAGAAACAGCTATGGAACAAATGAGAGTT 381

Qy 484 CCCTGAATGAAAAACACTCAAGGACTTACTTCTATAC 522
Db 382 CCCTGAATGAAAAACACTCAAGGACTTACTTCTATNC 420

RESULT 10

US-10-106-698-2133
; Sequence 2133, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA00591
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 2133
; LENGTH: 317
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (260)..(260)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc.feature
; LOCATION: (268)..(268)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc.feature
; LOCATION: (293)..(293)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-106-698-2133

Query Match 4.3%; Score 286; DB 15; Length 317;
Best Local Similarity 95.6%; Pred. No. 3.4e-49;
Matches 303; Conservative 0; Mismatches 13; Indels 1; Gaps 1;
Qy 1609 GAAGATAGATCCTGACCAAGAGTCAAGGAGGCTTCTGGCACTCATGTCCTCCAGAGCTTCTG 1728
Db 2 GAAGATAGATCCTGACCAAGAGTCAAGGAGGCTTCTGGCACTCATGTCCTCCAGAGCTTCTG 121
Qy 1669 GTTCTACAGTGTAGAACTCAAGAGGAGGCTTCTGGCACTCATGTCCTCCAGAGCTTCTG 1728
Db 62 GTTCTACAGTGTAGAACTCAAGAGGAGGCTTCTGGCACTCATGTCCTCCAGAGCTTCTG 121
Qy 1729 ATTGTTGTGAAGGTTTATTATATCTTTTCAAGCAATGATGATGGGCAAGACTTAG 1788
Db 122 ATTGTTGTGAAGGTTTATTATATCTTTTCAAGCAATGATGATGGGCAAGACTTAG 181
Qy 1789 ATTACTTTAATGATGAGCGGCAAAAGTGGCCCACTAATTAGTCTGACTG 1848
Db 182 ATTACTTTAATGATGAGCGGCAAAAGTGGCCCACTAATTAGTCTGACTG 240
Qy 1849 ATGCCCTTCTGACAGAACAGTATCTTCAGACCCTAATCAATAAGTCTTTTGAAGAAATG 1908
Db 241 ATGCCCTTCTGACAGAACAGTATCTTCAGACCCTAATCAATAAGTCTTTTGAAGAAATG 300
Qy 1909 TAAATGACTCTAAATCG 1925
Db 301 TAAATGACTCTAAATCG 317

RESULT 11

US-09-960-352-13367
; Sequence 13367, Application US/09960352
; Patent No. US20020137139A1

GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 13367
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 57-LIB3058-036-Q1-K1-G2
US-09-960-352-13367

Query Match 3.0%; Score 199.6; DB 9; Length 393;

Best Local Similarity 81.4%; Pred. No. 3.4e-31;
Matches 311; Conservative 0; Mismatches 49; Indels 22; Gaps 6;

Qy 5111 TTTCTTTTAAAGAAATTTATAGCATTTACTGTGTATTATTTAAATGCT-AAGCCAAAGTATCT 5169
Db 22 TGTCTTTATAGAAATTTTGTAGCATTTACTGTGTGTATTAAATGCTAAAGCCAAAGTATCT 81
Qy 5170 GCATTAGGTATACCTCTTTATGCCAATATGATTTTATGAGAGGCTCTTTTCAGATGTA 5229
Db 82 GCATT-----TTTATGCCAAGAACGATTTTAGTGAAGGCTCTTTTCAGATGTA 129
Qy 5230 ACCTATGAGGAATATCTGTTTGTATATGCCAGTTAGAACTACTGTTTCTAAAGT 5289
Db 130 ACCTATGAGGGAATATCTGTTTGTATATGCCAGTTAAATTTACTGTTTCTAAAGT 189
Qy 5290 CTG-TCAAAATTGTATTTCAGTGGCACAAACAGTTTTCAGGCTTTAGCTT-----AT 5343
Db 190 CTGTTAAAAATGTATTTCATGTCAC--AGACCAGTTTTCAGATCTTACACTTATATGAT 247
Qy 5344 AATCTTTGAATAAACTGA-TAATCTATTGTATTAATGGAGTGGAGACCTACCTCCAT 5402
Db 248 AATCTTTGAATAAAGCAGATTAGCTTATTGTTTAAATGGAGTGGAGACCTACCTCCAT 307
Qy 5403 AATTAGATAAATCTTTTGGATTATAATCAGAAATTTGCCTTTTTCTTCTCAAAATTAT 5462
Db 308 GATTAGATAAAGATTATTGCAATTAATCAAAATCAGAAATTTGCCTTTTTCTTCTCAAAATTAT 367
Qy 5463 TACATATGATGATTATATATAT 5484
Db 368 AATTAAATATATATATATATAT 389

RESULT 12

US-10-040-739-971
; Sequence 971, Application US/10040739
; Publication No. US20020173635A1
; GENERAL INFORMATION:

; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS

NUMBER OF SEQUENCES: 1519

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/10/040,739
APPLICATION NUMBER: US/10/040,739
FILING DATE: 07-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/036,520
FILING DATE: 03-JUN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 971:
SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 971:
US-10-040-739-971

Query Match 2.7%; Score 181.6; DB 14; Length 289;
Best Local Similarity 93.1%; Pred. No. 1.6e-27;
Matches 190; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2791 CTTGCACTCTCAGCACTTAACACACAGGTTTGAAGACTATGTTCCAAAGACAGA 2850
DB 18 CTTAGCACTCTCAGCACTTAACACACAGGTTTGAAGACTATGTTCCAAAGACAGA 77
QY 2851 AGAGAGTATGTTTGCAGATGTTATATGCCCAATGGTGAAGTTGCAGATACACAAAT 2910
DB 78 AGAGAGTATGTTTGCAGATGTTATATGCCCAATGGTGAAGTTGCAGATACACAAAT 137
QY 2911 TATCATCTGGAAGTAAAGATCTTCTGAAGACTTTAGTCTCTCTCCTCCTGATGCGCTA 2970
DB 138 TATCATCTGGAAGTAAAGATCTTCTGAAGACTTTAGTCTCTCTCCTCCTGATGCGCTA 197
QY 2971 TCACAGTAAACACAGTGGATCAT 2994
DB 198 TGTAGGAAATCAAGAAATCTT 221.

RESULT 13
US-09-864-761-13497
; Sequence 13497, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13497
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC018764.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4
US-09-864-761-13497

Query Match 2.6%; Score 172; DB 9; Length 556;
Best Local Similarity 100.0%; Pred. No. 2.3e-25;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4540 AGTGTTCTACTTTCTAAAGGACCGAGATTATCTATTTTATCAACTTCTTATCAGTTTG 4599
DB 208 AGTGTTCTACTTTCTAAAGGACCGAGATTATCTATTTTATCAACTTCTTATCAGTTTG 267
QY 4600 CAAAGAAATAGCCATGGCTTGTAGTCTGCCCTGCGCTCAGCTGAAACTCTTAAAA 4659
DB 268 CAAAGAAATAGCCATGGCTTGTAGTCTGCCCTCAGCTGAAACTCTTAAAA 327
QY 4660 GTAATGGATGAATAAATTGGACTCAGAGTTTCCATTGACACTGATATGTT 4711
DB 328 GTAATGGATGAATAAATTGGACTCAGAGTTTCCATTGACACTGATATGTT 379

RESULT 14
US-09-864-761-30043
; Sequence 30043, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 30043
; LENGTH: 168
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC018764.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4
; OTHER INFORMATION: NT HIT: g17662047, EVALUATE 5.00e-90
; OTHER INFORMATION: SWISSPROT HIT: Q05071, EVALUATE 8.80e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE883109.1, EVALUATE 7.00e-90
US-09-864-761-30043

Query Match 2.5%; Score 168; DB 9; Length 168;
Best Local Similarity 100.0%; Pred. No. 7.6e-25;
Matches 168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4542 GTGTTCTACTTTCTAAAGGACCCAGGATTTATCTATTTTATCAACTTCTTATCAGTTTGA 4601
DB 1 GTGTTCTACTTTCTAAAGGACCCAGGATTTATCTATTTTATCAACTTCTTATCAGTTTGA 60
QY 4602 AAAGAAATAGCCATGGCTTGTAGTCTGGCTGTGCTCCCTCACCTGAAAACTCTAAAGAAT 4661
DB 61 AAAGAAATAGCCATGGCTTGTAGTCTGGCTGTGCTCCCTCACCTGAAAACTCTAAAGAAT 120
QY 4662 AATGGATGATATAAATTTGGACTCAGATTTCCATTCAGACTGATG 4709
DB 121 AATGGATGATATAAATTTGGACTCAGATTTCCATTCAGACTGATG 168

RESULT 15
US-09-880-107-5/c
; Sequence 5, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107

; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 387
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA001902
; NAME/KEY: unsure
; LOCATION: (1)..(387)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-5

Query Match 1.9%; Score 124; DB 9; Length 387;
Best Local Similarity 96.9%; Pred. No. 1.7e-15;
Matches 157; Conservative 0; Mismatches 2; Indels 3; Gaps 3;
QY 6474 GTATGTATTCTG-CCATGTAAGTAATT-GAAACAGTCTTAAATAA-CCAAATGGTAGAGG 6530
DB 381 GTATGTATTCTGCCCATGTAGTAATTGGAACAGTCTTAAATAA-CCAAATGGTAGAGG 322
QY 6531 GCTGTTCCATGATGGGACAGCTTTGGATTGTTTTCATATAAATCTCTACATTCATAATAA 6590
DB 321 GCTGTTCCATGATGGGACAGCTTTGGATTGTTTTCATATAAATCTCTACATTCATAATAA 262
QY 6591 ATTGGAATTATGTGCTGAGTTTGGAGGCACATTTTGAAGT 6632
DB 261 ATTGGAATTATGTGCTGAGTTTGGAGGCACATTTTGAAGT 220

Search completed: May 18, 2004, 14:34:39
Job time : 1714.6 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 09:54:15 ; Search time 12011.2 Seconds
(without alignments)
17461.734 Million cell updates/sec

Title: US-09-744-167-1
Perfect score: 4839
Sequence: 1 gcaatcgaatcagcggac.....atcaaaaaaaaaaaaaaaa 4839

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: , 5940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.btg.*
- 3: gb.in.*
- 4: gb.in.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.btg.hum.*
- 31: em.btg.inv.*
- 32: em.btg.other.*
- 33: em.btg.mus.*
- 34: em.btg.pln.*
- 35: em.btg.rod.*
- 36: em.btg.mam.*
- 37: em.btg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4839	100.0	4839	9	AF104304	AF104304 Homo sapi
2	4679.8	96.7	4769	9	BC032680	BC032680 Homo sapi
3	4309.2	89.1	4565	9	AF130419	AF130419 Homo sapi
4	2397.6	49.5	2625	9	AF130420	AF130420 Homo sapi
5	2110.4	43.6	133799	9	AC105754	AC105754 Homo sapi
6	2110.4	43.6	150129	2	AC026909	AC026909 Homo sapi
7	1977	40.9	1977	6	AX809313	AX809313 Sequence
8	1552.2	32.1	89573	9	AL591398	AL591398 Human DNA
9	1545.8	31.9	4823	5	AF104305	AF104305 Xenopus l
C 10	1421.6	29.4	85802	10	AL671909	AL671909 Mouse DNA
C 11	1399.2	28.9	258787	2	AC095988	AC095988 Rattus no
C 12	1265.6	26.2	201330	2	AC118106	AC118106 Rattus no
C 13	898.8	18.6	1237	10	BC046341	BC046341 Mus muscu
14	875.4	18.1	164706	9	AC068614	AC068614 Homo sapi
15	740.8	15.3	1144	10	BC038255	BC038255 Mus muscu
16	699.4	14.5	888	6	AX677329	AX677329 Sequence
17	658	13.6	887	10	BC058416	BC058416 Mus muscu
18	576	11.9	576	6	AX677675	AX677675 Sequence
19	575.4	11.9	107627	9	AL513218	AL513218 Human DNA
C 20	575.4	11.9	123170	2	AL138792	AL138792 Homo sapi
C 21	575.2	11.9	192533	9	AC055876	AC055876 Homo sapi
C 22	564.8	11.7	151800	9	AC138649	AC138649 Homo sapi
C 23	564.6	11.7	178257	2	AC024350	AC024350 Homo sapi
24	564.4	11.7	3793	9	AK127003	AK127003 Homo sapi
25	562.2	11.6	6280	9	HSN805726	BS37424 Homo sapi
26	560.6	11.6	6632	9	AB002303	AB002303 Human mRN
27	559	11.6	5841	9	HSN804398	AL833087 Homo sapi
28	558.2	11.5	180741	9	AC091565	AC091565 Homo sapi
29	556.2	11.5	4620	9	AF434817	AF434817 Homo sapi
30	551.4	11.4	5438	10	AK122241	AK122241 Mus muscu
31	551.4	11.4	5480	10	BC042669	BC042669 Mus muscu
C 32	539	11.1	228470	2	AC136776	AC136776 Homo sapi
C 33	340	7.0	179370	10	AL626783	AL626783 Mouse DNA
34	339.6	7.0	5158	9	HSN809324	AX186602 Sequence
35	318	6.6	332	6	AX918602	BD054135 Sequence
36	318	6.6	332	6	ED054135	BD054135 Sequence
37	313.4	6.5	193289	2	AC127065	AC127065 Rattus no
38	245.8	5.1	4598	3	AF239997	AF239997 Drosophill
39	245.8	5.1	4651	3	DME310804	AJ310804 Drosophill
40	244.2	5.0	4660	3	AY051821	AY051821 Drosophill
C 41	202	4.2	54186	2	AC019894	AC019894 Drosophill
C 42	202	4.2	67833	3	AC004564	AC004564 Drosophill
43	202	4.2	179139	3	AC099307	AC099307 Drosophill
44	202	4.2	313634	3	AE003454	AE003454 Drosophill
45	190	3.9	4499	9	BC032227	BC032227 Homo sapi

ALIGNMENTS

RESULT 1
AF104304
LOCUS AF104304 4839 bp mRNA linear PRI 03-JAN-1999
DEFINITION Homo sapiens Smad anchor for receptor activation (SARA) mRNA,
complete cds.
ACCESSION AF104304
VERSION AF104304.1 GI:4092766
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4839)
AUTHORS Tsukazaki,T., Chiang,T.A., Davison,A.F., Attisano,L. and Wrana,J.L.
TITLE SARA, a FIVE domain protein that recruits Smad2 to the TGFbeta

receptor
Cell 95 (6), 779-791 (1998)
JOURNAL
MEDLINE
99081294
PUBMED
9865696
REFERENCE
2 (bases 1 to 4839)
Wana, J.L.
AUTHORS
Direct Submission
TITLE
Submitted (04-NOV-1998) Program in Developmental Biology, Hospital
JOURNAL
for Sick Children, 555 University Avenue, Toronto, Ontario M6G 1X8,
Canada

FEATURES

Location/Qualifiers

1..4839
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_type="HepG2"
1..4839
/gene="SARA"
439..4410
/gene="SARA"
/codon_start=1
/product="Smad anchor for receptor activation"
/protein_id="AAC99462.1"
/db_xref="GI:4092767"
/translation="MWIDENAVEDOLIKRNVNDQCSAVEVEGKCKGNLACLPEDEK
NVLVAVMHCNCKRLQNDLQDCNNYSQSLMDFSSLDNENRQTPQSFNSINESTE
KDMSEKQNDPLNRKTEGRSVNHLCPYSSDLSAVCSFQOLKODGSGIRDPNSAIT
SLTVDVSISSQGTDCPAVKQENYIPEDLTGKISSPRTDLGSPNFSHMSGEILMK
KEPABESTESLSRSLGKPLLLKPDMPNGSGRNDRCRSCDCLVENEVRADENEGEHE
ETLGTTEFLNTEHFSQSDMTNKLTKLNEINDSQVNEKEKEFLQISQPEDTNGSG
GQCVLADAGLDLGTCTISESECEPSTVIDTPAANYLSNGCDSDYGMQDPGVSPVPT
LPSKEDSYTEKEIEESKSECVSNLYEORGNATEGSGLLNLSGDLMKKNLYLHNFCS
QVPSVLGSSPKYVASLPSISVPPGGARPQKPSNLKQIPKPLSDHLQNDPFPANSGNN
TKNKDILIGKNGSNATNVSFSLNINVDINGHLESIEAISTRLCLALAPDS
PNDLRAGQGISARKPFTLLGEVAPVSDPSQPNCKVCEARFTFKRHLRACRCK
VFCASCCGLKGLKLYRKEARVCVICHVILNMAQWENMMSSQSPNPNNAEYGS
TIPPLQQAQAGALSPPPTVPVGLKHPGEAQAQFQRRVWFPADGILPNEVAD
AAKLTMTNSTAGTLAVSHDPVKVPTTSPLEAETDIDCLFSGITQVSPVGSANLIP
EDGLPILITSGVGDVAVERKPSQISVMQOLEDGDPPLVFLVNLNLSMKLYNVY
NRKWCFTITKGNHVAQSEIIVLLOQCLPDEKLPKIDFENFVQLYRDLALGNVSNL
HSFSGSLGSKHGGLFVSTYQSLQDLVLPPLFGLILQKWTETPKVPPFPL
MLRGAELRYLPCLFVSFRKPLFGETHTIMNLADFPNYQYTLFVQGLVDMVEV
RKTGIKIPSNRYNEMKAMNKHNEHLVLAGCFNEKADSHLVCVQNDGNYQTAISI
HNQPKVTGAGFFVFSGLKSSGVLAKSSIVEDGVVQVTAENMDLSRLALREKDF
TITCKDADEPPOEHIHTQWVDDDKNVSKGVSPIDGKSMETITNVI PHGSEYKANG
KVIKRWTEVFLENDHNLCLSDPADHSLRTEHVAKAFCLALCPHLKLLKEDGNTKGL
RVTILSDQVQVQAGSNGQPLPSQYMNLDLSNLVPIHGGACQJSEGGPVNELLFYILE
NIV"

gene

CDS

Qy	301	AATGAATCTGCAGTTCTTAATGAGTCAACACCACTGAAAGTCTTCTCCCTGGCTCAT	360
Db	301	AATGAATCTGCAGTTCTTAATGAGTCAACACCACTGAAAGTCTTCTCCCTGGCTCAT	360
Qy	361	TCAGTCTCCCTGACCAACAGAGGAAGAGTCACTGTGCTAATGGAAGAGTCAATCTA	420
Db	361	TCAGTCTCCCTGACCAACAGAGGAAGAGTCACTGTGCTAATGGAAGAGTCAATCTA	420
Qy	421	AATCCAGAGATTGCCCAATGTGATGAAATGCTGTGTCAGAACACCACTTAAT	480
Db	421	AATCCAGAGATTGCCCAATGTGATGAAATGCTGTGTCAGAACACCACTTAAT	480
Qy	481	AAGAGAACTATAGTTGGGATGATCAATGAGTGTGTTGAAAGTGGGAGAGAAATGT	540
Db	481	AAGAGAACTATAGTTGGGATGATCAATGAGTGTGTTGAAAGTGGGAGAGAAATGT	540
Qy	541	GGAACTCTGCTGCTGCCAGATGAGAGAAATGTTCTTTGTTGAGCGTCAATGATAC	600
Db	541	GGAACTCTGCTGCTGCCAGATGAGAGAAATGTTCTTTGTTGAGCGTCAATGATAC	600
Qy	601	TGTGATAAAGGACATTACAAAACGATTTACAGGATTTGTAATAATATATAGTCAATCC	660
Db	601	TGTGATAAAGGACATTACAAAACGATTTACAGGATTTGTAATAATATATAGTCAATCC	660
Qy	661	CTTATGATGCTTTAGCTGTTCACTGGATTAATGAAAACAGACAACTGATCAATTTAGT	720
Db	661	CTTATGATGCTTTAGCTGTTCACTGGATTAATGAAAACAGACAACTGATCAATTTAGT	720
Qy	721	TTTAGTATAAATGAGTCCACTCAAAAAGATATGAATTCAGAGAAAACAAATGATCCATG	780
Db	721	TTTAGTATAAATGAGTCCACTCAAAAAGATATGAATTCAGAGAAAACAAATGATCCATG	780
Qy	781	AATAGACCGAAACAGAGGGAGATCTGTAAACATCTGTGCTCTACTCTCTCTGATAGT	840
Db	781	AATAGACCGAAACAGAGGGAGATCTGTAAACATCTGTGCTCTACTCTCTCTGATAGT	840
Qy	841	CTAGCCAGTGTCTGTCCCTCTTCAAAATTAAGAGATGAGGAGATATAGTAGAGACCC	900
Db	841	CTAGCCAGTGTCTGTCCCTCTTCAAAATTAAGAGATGAGGAGATATAGTAGAGACCC	900
Qy	901	TCCATGCTCTGCATTACAAAGTTTAAACGGTTGATTCAGTAATCTCATCCAGGGAACAGAT	960
Db	901	TCCATGCTCTGCATTACAAAGTTTAAACGGTTGATTCAGTAATCTCATCCAGGGAACAGAT	960
Qy	961	GGATGCTCTGCTGTTAAAGCAAGAGACTATATACAGATGAGGAGCTCACTGGCAAA	1020
Db	961	GGATGCTCTGCTGTTAAAGCAAGAGACTATATACAGATGAGGAGCTCACTGGCAAA	1020
Qy	1021	ATCAGCTCTCTAGGACAGATCTAGGGAGTCCAAATTCCTTTCCACATGAGTGAAGGG	1080
Db	1021	ATCAGCTCTCTAGGACAGATCTAGGGAGTCCAAATTCCTTTCCACATGAGTGAAGGG	1080
Qy	1081	ATTGATGAAAAAGAGCCAGAGAGAGACACCTGAAAGATCCCTCCGCTCTGGT	1140
Db	1081	ATTGATGAAAAAGAGCCAGAGAGAGACACCTGAAAGATCCCTCCGCTCTGGT	1140
Qy	1141	TTACCTTCTCTCTCAAAACAGACATGCTTAATGGGTCTGGAAGGAATAATGACTGAA	1200
Db	1141	TTACCTTCTCTCTCAAAACAGACATGCTTAATGGGTCTGGAAGGAATAATGACTGAA	1200
Qy	1201	CGGTGTTTCAGATGCTGCTGCTTAATGAAGTTAGGCTGATGAAAAATGAAGTTATGAA	1260
Db	1201	CGGTGTTTCAGATGCTGCTGCTTAATGAAGTTAGGCTGATGAAAAATGAAGTTATGAA	1260
Qy	1261	CATGAAGAACTCTTGGCACTACAGAAATTCCTTAATATGACAGAGCAATTTCTCTGAATCT	1320
Db	1261	CATGAAGAACTCTTGGCACTACAGAAATTCCTTAATATGACAGAGCAATTTCTCTGAATCT	1320
Qy	1321	CAGGACATGACTTAATGGAAGTTGACTAACTAAATGAGATGAATGATAGCCAAAGTAAAC	1380
Db	1321	CAGGACATGACTTAATGGAAGTTGACTAACTAAATGAGATGAATGATAGCCAAAGTAAAC	1380

ORIGIN

Query Match	100.0%	Score 4839;	DB 9;	Length 4839;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 4839;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GCATCTGAATCAGCAGACTGGCTGTGGTGGTGGCAGCAGATCATGATGATGACCCGAGA	60	
Db	1	GCATCTGAATCAGCAGACTGGCTGTGGTGGTGGCAGCAGATCATGATGATGACCCGAGA	60	
Qy	61	AGTCTGTTCTTATCACCTGTGTAAGGGGAAAAAGTTTAAACCAAGTCTCTTAAGTGGTG	120	
Db	61	AGTCTGTTCTTATCACCTGTGTAAGGGGAAAAAGTTTAAACCAAGTCTCTTAAGTGGTG	120	
Qy	121	TTTCTCTACCGATGAGAAATTAATTCACAGCAGAGCTTACAACTCGGACAAAGTGTTA	180	
Db	121	TTTCTCTACCGATGAGAAATTAATTCACAGCAGAGCTTACAACTCGGACAAAGTGTTA	180	
Qy	181	GATGAATTTGACAAAACGAGATGAAACAGTTTCTCTTCTTATTTGGATACAAAGTGG	240	
Db	181	GATGAATTTGACAAAACGAGATGAAACAGTTTCTCTTCTTATTTGGATACAAAGTGG	240	
Qy	241	AATAAGATTTAGATCCCTCTTCTCACGGGTGTGCTTAATTAACCTACTTTGGCCAGTGTG	300	
Db	241	AATAAGATTTAGATCCCTCTTCTCACGGGTGTGCTTAATTAACCTACTTTGGCCAGTGTG	300	

QY 1381 GAAGAAAAGGAAAGTTTCTACAGATTAGTCAGCCTGAGGACACTAATGCTGATAGTGA 1440
DB 1381 GAAGAAAAGGAAAGTTTCTACAGATTAGTCAGCCTGAGGACACTAATGCTGATAGTGA 1440
QY 1441 GGACAGTGTGTGGATTGGCAGATGCGGCTAGATTAAAGGAAGTTTGCATTAGTGA 1500
DB 1441 GGACAGTGTGTGGATTGGCAGATGCGGCTAGATTAAAGGAAGTTTGCATTAGTGA 1500
QY 1501 AGTGAAAGATTGTGATTTCTCCACTGTTATAGACACACCCAGCAGCAAAATATCTAT 1560
DB 1501 AGTGAAAGATTGTGATTTCTCCACTGTTATAGACACACCCAGCAGCAAAATATCTAT 1560
QY 1561 GGTGTGATTCCTATGAAATGCAAGACCCAGGTGTTTCTTTGTTCCAAAGACTTTACCC 1620
DB 1561 GGTGTGATTCCTATGAAATGCAAGACCCAGGTGTTTCTTTGTTCCAAAGACTTTACCC 1620
QY 1621 TCCAAAGAGATTTCAGTAACAGAGAAAGAAATAGAGAAAGCAAGTCAGAAATGCTAC 1680
DB 1621 TCCAAAGAGATTTCAGTAACAGAGAAAGAAATAGAGAAAGCAAGTCAGAAATGCTAC 1680
QY 1681 TCAAATATTTATGAACAGAGAGGAAATGAGGCCACAGAGGGAGTGGAATCTTTTAAAC 1740
DB 1681 TCAAATATTTATGAACAGAGAGGAAATGAGGCCACAGAGGGAGTGGAATCTTTTAAAC 1740
QY 1741 AGCACTGGTGACCTTAATGAAGAAAATTTATCATTAATTTCTAGTCAAGTTCCATCA 1800
DB 1741 AGCACTGGTGACCTTAATGAAGAAAATTTATCATTAATTTCTAGTCAAGTTCCATCA 1800
QY 1801 GTGCTTGGGCAATCTTCCCCAAGGTAGTAGCAAGCTGCCATCTATCAGTGTTCCTTTT 1860
DB 1801 GTGCTTGGGCAATCTTCCCCAAGGTAGTAGCAAGCTGCCATCTATCAGTGTTCCTTTT 1860
QY 1861 GGTGGTGCAAGCCCAAGCAACCTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1920
DB 1861 GGTGGTGCAAGCCCAAGCAACCTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1920
QY 1921 GACCAATTTACAAAATGACTTTCTCGCAACAGTGGAAATTAATCTAATAAATAAATAAT 1980
DB 1921 GACCAATTTACAAAATGACTTTCTCGCAACAGTGGAAATTAATCTAATAAATAAATAAT 1980
QY 1981 ATTCCTGGGAAAGCAAAATTAGGGGAAATTCAGCAACCAATGATGAGTCCATCTTTG 2040
DB 1981 ATTCCTGGGAAAGCAAAATTAGGGGAAATTCAGCAACCAATGATGAGTCCATCTTTG 2040
QY 2041 GGAAACATCTTAATGTGATACAAATGGGAAACATTTAGAAAGTTATGAGGCTGAGATC 2100
DB 2041 GGAAACATCTTAATGTGATACAAATGGGAAACATTTAGAAAGTTATGAGGCTGAGATC 2100
QY 2101 TCCACTAGACCATGCTTGCATTAGCTCCAGATAGCCAGATAATGATCTCAGAGCTGGT 2160
DB 2101 TCCACTAGACCATGCTTGCATTAGCTCCAGATAGCCAGATAATGATCTCAGAGCTGGT 2160
QY 2161 CAGTTTGGAAATTTCTGCCAGAAAGCCATTACACAGCTGGTGAGTGGCTCCAGTATGG 2220
DB 2161 CAGTTTGGAAATTTCTGCCAGAAAGCCATTACACAGCTGGTGAGTGGCTCCAGTATGG 2220
QY 2221 GTACCGGATTTCTCAGGCTCCAAATTTGCATGAAATGTGAAGCAGGTTTACATTCACCAA 2280
DB 2221 GTACCGGATTTCTCAGGCTCCAAATTTGCATGAAATGTGAAGCAGGTTTACATTCACCAA 2280
QY 2281 AGGAGGCAATCACTGCAGAGCATGTGGGAAAGTTTCTGTGCTTCTGTGCTGAGCCTGAAA 2340
DB 2281 AGGAGGCAATCACTGCAGAGCATGTGGGAAAGTTTCTGTGCTTCTGTGCTGAGCCTGAAA 2340
QY 2341 TGTAAATGTTTATACATGCAGAGAAAGCAAGTGTGTAATCTGCCATTCAGTG 2400
DB 2341 TGTAAATGTTTATACATGCAGAGAAAGCAAGTGTGTAATCTGCCATTCAGTG 2400
QY 2401 CTAATGAATGCTCAAGCCTGGGAGAAACATGATGAGTGGCTCAAGCCAGAGCCCTAACCT 2460
DB 2401 CTAATGAATGCTCAAGCCTGGGAGAAACATGATGAGTGGCTCAAGCCAGAGCCCTAACCT 2460
QY 2461 AACAAATCTGCTGGAATATCTGTTCTACTATCCCTTCCCTTGCAGCAAGCTCAGGCTCAGGA 2520

DB 2461 AACAAATCTGCTGGAATATCTGTTCTACTATCCCTTGCAGCAAGCTCAGGCTCAGGA 2520
QY 2521 GCTCTGAGCTCTCCACCTCCCACTCTGATGTGTACCTGTGGGAGTTTAAAGCACCTCGA 2580
DB 2521 GCTCTGAGCTCTCCACCTCCCACTCTGATGTGTACCTGTGGGAGTTTAAAGCACCTCGA 2580
QY 2581 GCAGAAAGTGTGTGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
DB 2581 GCAGAAAGTGTGTGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
QY 2641 AATGAGAAAGTGTGTGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
DB 2641 AATGAGAAAGTGTGTGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
QY 2701 CTGGCTGTGTACACGAG 2760
DB 2701 CTGGCTGTGTACACGAG 2760
QY 2761 GATATTTGCTATTTCTCTGGGAGTAACTCAGGTTGGAAGTCTGTGGAGTGCATATG 2820
DB 2761 GATATTTGCTATTTCTCTGGGAGTAACTCAGGTTGGAAGTCTGTGGAGTGCATATG 2820
QY 2821 AATCTTATTTCTGAAAGTGGCCTTCTCCATTTCTCATCTCCACTGGTGTAAAAGAGAGAC 2880
DB 2821 AATCTTATTTCTGAAAGTGGCCTTCTCCATTTCTCATCTCCACTGGTGTAAAAGAGAGAC 2880
QY 2881 TATGCTGTGGAG 2940
DB 2881 TATGCTGTGGAG 2940
QY 2941 CTTGACCCACTTGTATTTGTTTAAATGCAAAATTTGTTGTCAATGTTAAAAATTTGTAAT 3000
DB 2941 CTTGACCCACTTGTATTTGTTTAAATGCAAAATTTGTTGTCAATGTTAAAAATTTGTAAT 3000
QY 3001 TATGTGAACAG 3060
DB 3001 TATGTGAACAG 3060
QY 3061 GAGATAGTCAATCTTCTACAGTGTTCCTGGAGTAAAGACATGCTGATTTATATATG 3120
DB 3061 GAGATAGTCAATCTTCTACAGTGTTCCTGGAGTAAAGACATGCTGATTTATATATG 3120
QY 3121 AATCACTTTGTGAG 3180
DB 3121 AATCACTTTGTGAG 3180
QY 3181 CATTCCTTCTCAGTCAAAAGTTTCTTGGAGTAAAGACATGCTGATTTATATATG 3240
DB 3181 CATTCCTTCTCAGTCAAAAGTTTCTTGGAGTAAAGACATGCTGATTTATATATG 3240
QY 3241 ACATCTACCTACAGTCAAG 3300
DB 3241 ACATCTACCTACAGTCAAG 3300
QY 3301 ATTCTTATCCAGAAATGGGAAACTCCTGGGCTAAAGTATTTCCATCCTGCTGATGTTG 3360
DB 3301 ATTCTTATCCAGAAATGGGAAACTCCTGGGCTAAAGTATTTCCATCCTGCTGATGTTG 3360
QY 3361 AGACTTGGAGCTGAATATGACTTTTATCATGCCCACTATTCAGTGTCAAGATTTCCGAAAG 3420
DB 3361 AGACTTGGAGCTGAATATGACTTTTATCATGCCCACTATTCAGTGTCAAGATTTCCGAAAG 3420
QY 3421 CCATTTGTTGGAG 3480
DB 3421 CCATTTGTTGGAG 3480
QY 3481 CAGTATACCTTGCAGTGTTCAGGTTTGGTGTGATGATGAGTGTTCGGAAGCTAGC 3540
DB 3481 CAGTATACCTTGCAGTGTTCAGGTTTGGTGTGATGATGAGTGTTCGGAAGCTAGC 3540
QY 3541 ATCAAATTTCCCAAGCAACAGATACAAATGAGATGATGAAAGCCATGAAGTCCAAATGAG 3600

3541 ATCAAAATTCCTCCAGCAACAGATACAAATGAGATGATGAAGCCATGAACCAAGTCCATGAG 3600
 3601 CATGCTCTGGAGAGAGGCTGCTGCTTCAATGAAGAGGAGAGCTCTCATCTTGTGTGTGTA 3660
 3601 CATGCTCTGGAGAGAGGCTGCTGCTTCAATGAAGAGGAGAGCTCTCATCTTGTGTGTGTA 3660
 3661 CAGATGATGATGGAAGAACTATCAGACCCAGGCTATCAGTATTCAATCAGCCAGAGAAA 3720
 3661 CAGATGATGATGGAAGAACTATCAGACCCAGGCTATCAGTATTCAATCAGCCAGAGAAA 3720
 3721 GTGACTGGTGCAGTTCTTGTGTGTCAGTGGCGCTTGAATCTCTTCTTGGTACCTT 3780
 3721 GTGACTGGTGCAGTTCTTGTGTGTCAGTGGCGCTTGAATCTCTTCTTGGTACCTT 3780
 3781 GCCAAGTCCAGTATTGTGGAGATGGTGTATGGTCCAGATTACTGCAGAGAAATGGAT 3840
 3781 GCCAAGTCCAGTATTGTGGAGATGGTGTATGGTCCAGATTACTGCAGAGAAATGGAT 3840
 3841 TCCITGAGCAGGCACTGCGAGATGAAGGACTTCCACATCAGCTGTGGAGGCGGAC 3900
 3841 TCCITGAGCAGGCACTGCGAGATGAAGGACTTCCACATCAGCTGTGGAGGCGGAC 3900
 3901 GCGAGGAGACCCAGAGACATCCACATCCAGTGGTGGATGATGACAGAACTGTAGC 3960
 3901 GCGAGGAGACCCAGAGACATCCACATCCAGTGGTGGATGATGACAGAACTGTAGC 3960
 3961 AAGGCTGTCTAAGTCTTATAGATGGGAGTCCATGGAGACTATAACAAATGTGAAGATA 4020
 3961 AAGGCTGTCTAAGTCTTATAGATGGGAGTCCATGGAGACTATAACAAATGTGAAGATA 4020
 4021 TTCCATGGATCAGATATAAAGCAAAATGGAAGTAATCAGATGAGACAGAGTGTCTTTC 4080
 4021 TTCCATGGATCAGATATAAAGCAAAATGGAAGTAATCAGATGAGACAGAGTGTCTTTC 4080
 4081 CTAGAAAACAGATGACACAGCAATGCTCAGTGTCTGATCTGAGATCAGATGAGATGACT 4140
 4081 CTAGAAAACAGATGACACAGCAATGCTCAGTGTCTGATCTGAGATCAGATGAGATGACT 4140
 4141 GAGCATGTTGCAAAAGCTTTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4200
 4141 GAGCATGTTGCAAAAGCTTTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4200
 4201 GATGAAGTACCAACTGGGACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4260
 4201 GATGAAGTACCAACTGGGACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4260
 4261 GCAGGAGCAATGGCGAGCCCTTCCCTCGCAGTACATGATGATGATGATGATGATGATGATG 4320
 4261 GCAGGAGCAATGGCGAGCCCTTCCCTCGCAGTACATGATGATGATGATGATGATGATGATG 4320
 4321 GTGCGGTGATCCATGGAGGGGCTGCGAGCTTAGTGAGGGCCCGCTTGTCTGATGAGAACTC 4380
 4321 GTGCGGTGATCCATGGAGGGGCTGCGAGCTTAGTGAGGGCCCGCTTGTCTGATGAGAACTC 4380
 4381 ATCTTTTATTTCTGGAAGAACTCGTATAACAGAGAGAGCTTCATTTTCTGTTTCAG 4440
 4381 ATCTTTTATTTCTGGAAGAACTCGTATAACAGAGAGAGCTTCATTTTCTGTTTCAG 4440
 4441 ACTTGTTCGACAGCAGTATACCAATATTTGCACTTTTAAACTTGAAGATTAAGCT 4500
 4441 ACTTGTTCGACAGCAGTATACCAATATTTGCACTTTTAAACTTGAAGATTAAGCT 4500
 4501 TTTGTTTACACTTATTAATGGGGTGGGAATAGGTTGGAGTGGGGTTTGGAGACGGGT 4560
 4501 TTTGTTTACACTTATTAATGGGGTGGGAATAGGTTGGAGTGGGGTTTGGAGACGGGT 4560
 4561 GGGAAAGGCTGTTGGGGGACCGATGTTCCATAATTTCTAAGTCTTCTATGCAATTTGCCA 4620
 4561 GGGAAAGGCTGTTGGGGGACCGATGTTCCATAATTTCTAAGTCTTCTATGCAATTTGCCA 4620
 4621 CCAAGAGATCTGGCAGCTTCTCTCTCTGACACAGAGTATGCTATCTCTGAGCTAAT 4680
 4621 CCAAGAGATCTGGCAGCTTCTCTCTCTGACACAGAGTATGCTATCTCTGAGCTAAT 4680

QY 4681 CCCCTTCTGTACTTGTAGACAGAAATTCGGTCTCTCTCTCAAGATTTTACTTATGGTCA 4740
 Db 4681 CCCCTTCTGTACTTGTAGACAGAAATTCGGTCTCTCTCTCAAGATTTTACTTATGGTCA 4740
 QY 4741 TGTGCTCAGAAATGCTCAAAATGGTGTCAACACCATCAACAGGGTGGATGGAGGGCAGAG 4800
 Db 4741 TGTGCTCAGAAATGCTCAAAATGGTGTCAACACCATCAACAGGGTGGATGGAGGGCAGAG 4800
 QY 4801 GGGAAATAAATATATAAGCATCAAAAAAATAAAAAA 4839
 Db 4801 GGGAAATAAATATATAAGCATCAAAAAAATAAAAAA 4839

RESULT 2
 LOCUS BC032680 4769 bp mRNA linear PRI 08-OCT-2003
 DEFINITION Homo sapiens MAD, mothers against decapentaplegic homolog (Drosophila) interacting protein, receptor activation anchor, mRNA (cDNA clone MGC:45259 IMAGE:585817), complete cds.
 ACCESSION BC032680
 VERSION BC032680.1 GI:21618607
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 MAMMALIA; Eutheria; Primates; Catarrhini, Hominoidea; Homo.

REFERENCE 1 (bases 1 to 4769)
 Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.W., Schuler G.D., Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S., Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E., Jones S.J. and Marra M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

TITLE
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16898-16903 (2002)
 MEDLINE 22388257
 PUBMED 12477932
 REFERENCE 2 (bases 1 to 4769)
 DIRECT SUBMISSION Strausberg R.
 TITLE
 JOURNAL Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
 COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@nigr.nih.gov
 Akhtar N., Ayale K., Beckstrom-Sternberg S.M., Benjamin B., Blakesley R.W., Bouffard G.G., Breen K., Brinkley C., Brooks S., Dietrich N.L., Granite S., Guan X., Gupta K., Haghigati P., Hansen N., Ho S.-L., Karlins E., Kwong P., Laric P., Legaspi R.,

Maduro, Q.L., Masiello, C., Maskeri, B., Mastriani, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
Series: IRAK Plate: 69 Row: p Column: 5
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4759059.

FEATURES

source

1..4769
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:45259 IMAGE:5585817"
/tissue_type="Testis, embryonal carcinoma"
/clone_lib="NIH MGC_92"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
1..4769
/gene="MADHIP"
/note="synonyms: SARA, ZFYVE9"
/db_xref="LocusID:9372"
/db_xref="MIM:603755"
12..4289
/codon_start=1
/product="MADHIP protein"
/protein_id="AAH32680.1"
/db_xref="GI:21618608"
/db_xref="LocusID:9372"
/translation="MENYFOAEYNLDKVLDEPONEDETVSSLTLDTKWNLDPDS
HRLSFTPLASVNSAVNSQPOLKVFSLAHSAPLITEEDHCANGQDCNLPRIAT
MWIDENVAEDQLIKRNYSDWDCSAVEGKKGKGLACLPDEKKNLVAVMKNCKR
TLQNDLQCNNSYQMLDAPSCSLDNENRQTFQSFINEKKNMSEKQMDPLNR
KPTGRSVNHLCPSTSLASVCSQSLKDDSGISDRDPSMSAITSLTVDVSISSQGT
GCPAVKQENYIPEDLTGKISSPRTDLGSPNFSHMSGILMKRPAEBSFTSESLR
SGPLLLKPMQPSGNNDCRSCDLVNEVRADENEGYEHEETLGTFFLNMTET
PSESQDMNKLTKLNEMNSQVNEKEKELQISQPEDTNGDSGGQGVGLADGLK
GTCISEEEDCFSTVIDTPAANYLSNGSDSGWDPGVFVPVKILSKESVTEBEKEI
EESKESYNIYERGEATFEGSGLLNSTGDLMMKNYLHNFCSQVPSVLGQSPKV
ASLPSISVPFGARPKFNSLKLQIPKLDHLQNDPANSNNKNNDILGAKLG
ENSNATYSPSLGNSVNTNGHELYEAEISIRCLALAPSPNDLRAGQFISA
RKPTTLGEVAPVNPDSQAPNCKCEARFTTKRHHCRACGVKSCCSLKKLL
YMDKEARVCVICHSLVMAQAWENMSASSQSPNPNPAEYCSIIPLQOASGAL
SSPPTVNPVGVVLKHPGAEVAQPREQRRVWEADGILPNGEVADAALKTWNGTSSAGT
LAVSHDPVPTVTSPLFAETDICIPLFGSITQVSPVGSANMLPEDGLPILISTGVK
GDYAVEKPSQISVYQLEDGDPPLVFLVNLNLSMKVLYNVRKWCFTTNGMHA
VGQSEIVILLQCLEDEKLPKDIENHFVQLYRDALAGNVVSNLGHFSFQSLGSKKH
GGFLYVTSYQSLQDLVLPYPLGILIQKWTFFWAKVFPRLMLRLGAEYRLYPCP
LPSVRFKPLFGETGHTIMNLLADFRNQVTLPVVQGLVDMVRKTSLKIPSNRYNE
MKAMNKNHVLGAGGACFEKADSHLVQVNDGNYQTOAISIHNPVKVTPGASPV
PSGALKSSGVYLAQSSIVEDGVWQITAEKNDLSRQALREMKOFTITCGKADAEBOE
HIHQVDDDDKNNKGVSPIDGKSMETITNKKIFPHGEYKANGKVIWTEVFELEND
DOHNLSDPADSHSLTHVAKFCLCPHLKLLKEDGMKGLRVLDSDDQVGYQAG
SNGQPLPSQYNNLDSALVPVHGGACQLSEGPVVMELIFYLENIY"
2091..2282

CDS

/note="FYVE: Region: FYVE zinc finger. The FYVE zinc
finger is named after four proteins that it has been found
in: Fab1, YOTB/ZK632.12, Vac1, and EEA1. The FYVE finger
has been shown to bind two Zn++ ions. The FYVE finger has
eight potential zinc coordinating cysteine positions. Many
members of this family also include two histidines in a
motif R-HHC+XCG, where + represents a charged residue and
X any residue. We have included members which do not
conserve these histidine residues but are clearly related"
/db_xref="CDD:pfam01363"

misc_feature

96.7%; Score 4679.8; DB 9; Length 4769;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4692; Conservative 2; Indels 1; Gaps 1;

ORIGIN

Query Match 96.7%; Score 4679.8; DB 9; Length 4769;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4692; Conservative 2; Indels 1; Gaps 1;

129 CCGATGGAGAAATTACTTCCAGCAGAGCTTACACCTGGGGACAAGGTGTGTAGATGAATT 188
9 CCGATGGAGAAATTACTTCCAGCAGAGCTTACACCT-GGACAAGGTGTGTAGATGAATT 67
189 TGAACAAAACGAGAGATGAACAGATTTCTTCTACTTTATTGGATACAAAGTGAATAAGAT 248
68 TGAACAAAACGAGAGATGAACAGATTTCTTCTACTTTATTGGATACAAAGTGAATAAGAT 127
249 TCTAGATCCCCCTTCTCACCGGCTGTCAATTTAAACCTACTTTGGCCAGTGTGAATGAAATC 308
128 TCTAGATCCCCCTTCTCACCGGCTGTCAATTTAAACCTACTTTGGCCAGTGTGAATGAAATC 187
309 TGCAGTTTCTAATAGTACACACCACTGAAGTCTTCTCCCTGGTCAATTCAGCTCC 368
188 TGCAGTTTCTAATAGTACACACCACTGAAGTCTTCTCCCTGGTCAATTCAGCTCC 247
369 CTTGACCAACAGAGAGAGGATCACTGTGTCTAATGGACAGGACTGTAAATCTAAATCCAGA 428
248 CTTGACCAACAGAGAGAGGATCACTGTGTCTAATGGACAGGACTGTAAATCTAAATCCAGA 307
429 GATTGCCAATGTGGATGAATAATGCTTTCGACAGACCACTTAATTAAGAGAA 488
308 GATTGCCAATGTGGATGAATAATGCTTTCGACAGACCACTTAATTAAGAGAA 367
489 CTATAGTTGGGATGATCAATGCACTGCTTGAAGTGGGAGAGAGAAATGTGGAACCT 548
368 CTATAGTTGGGATGATCAATGCACTGCTTGAAGTGGGAGAGAGAAATGTGGAACCT 427
549 GCGTTGCTGCAGATGAGAGAAATGTTCTTGTGTAGCCGTCATGCATACTGTGATAA 508
428 GCGTTGCTGCAGATGAGAGAAATGTTCTTGTGTAGCCGTCATGCATACTGTGATAA 487
509 AAGGACATTAACAAAGATTTACAGATTTAATAATTAATAGTCAATCCCTTATGGA 668
488 AAGGACATTAACAAAGATTTACAGATTTAATAATTAATAGTCAATCCCTTATGGA 547
669 TGTCTTTAGCTGTCTACTCGATAATGAACACAGACAACTGATCAATTTAGTTTAGTAT 728
548 TGTCTTTAGCTGTCTACTCGATAATGAACACAGACAACTGATCAATTTAGTTTAGTAT 607
729 AATAGTCTCACTGAACAAAGATTTAGATTCAGAGAACAAATGGATCCATTCGAATAGCC 788
608 AATAGTCTCACTGAACAAAGATTTAGATTCAGAGAACAAATGGATCCATTCGAATAGCC 667
789 GAAACACAGAGGGAGATCTGTAAACCATCTGTGTCTCTTCTCATCTGTAGTCTAGCCAG 848
668 GAAACACAGAGGGAGATCTGTAAACCATCTGTGTCTCTTCTCATCTGTAGTCTAGCCAG 727
849 TGTCTGTTCCCTTCAATTAAGGATGACGGAAGTATAGTAGAGACCCCTCCATGTC 908
728 TGTCTGTTCCCTTCAATTAAGGATGACGGAAGTATAGTAGAGACCCCTCCATGTC 787
909 TGCATTACAAGTTTAAACGGTTGATTCACTGTGTCTCTTCTCATCTGTAGTCTAGCCAG 968
788 TGCATTACAAGTTTAAACGGTTGATTCACTGTGTCTCTTCTCATCTGTAGTCTAGCCAG 847
969 TGCATTACAAGTTTAAACGGTTGATTCACTGTGTCTCTTCTCATCTGTAGTCTAGCCAG 1028
848 TGCATTACAAGTTTAAACGGTTGATTCACTGTGTCTCTTCTCATCTGTAGTCTAGCCAG 907
1029 TGCATTACAAGTTTAAACGGTTGATTCACTGTGTCTCTTCTCATCTGTAGTCTAGCCAG 1088
908 TGCATTACAAGTTTAAACGGTTGATTCACTGTGTCTCTTCTCATCTGTAGTCTAGCCAG 967
1089 GAAACACAGAGGGAGATCTGTAAACCATCTGTGTCTCTTCTCATCTGTAGTCTAGCCAG 1148
968 GAAACACAGAGGGAGATCTGTAAACCATCTGTGTCTCTTCTCATCTGTAGTCTAGCCAG 1027
1149 GCTTCTCAACACAGATGCTTAATCGGTCTGGAAGAAATATGACTGCAAGGTTGTC 1208
1028 GCTTCTCAACACAGATGCTTAATCGGTCTGGAAGAAATATGACTGCAAGGTTGTC 1087
1209 AGATTGCTTGTGCTTAAAGAGTTAGGCTGTAGTGAAGTGAAGTTTATGCAACATGAAGA 1268

QY 3429 TGGAGACGGGGCATACCATCATGAATCTTCTTGAGAGCTTCAGAAAATTACAGTATAC 3488
Db 3308 TGGAGACGGGGCATACCATCATGAATCTTCTTGAGAGCTTCAGAAAATTACAGTATAC 3367
QY 3489 CTTGCCAGTAGTCAAGGTTTGGTGGTATGATGGAAGTTTGGGAAACTAGCATCAAAAT 3548
Db 3368 CTTGCCAGTAGTCAAGGTTTGGTGGTATGATGGAAGTTTGGGAAACTAGCATCAAAAT 3427
QY 3549 TCCCAGCAACAGATACAAATGAGATGATGAAGCCATGAAACAAGTCCAATGAGCATGTCTT 3608
Db 3428 TCCCAGCAACAGATACAAATGAGATGATGAAGCCATGAAACAAGTCCAATGAGCATGTCTT 3487
QY 3609 GGCAGGAGTCCCTGCTTCAATGAAAGGAGAGACTCTCATCTTGTGTGTGTAAGAATGA 3668
Db 3488 GGCAGGAGTCCCTGCTTCAATGAAAGGAGAGACTCTCATCTTGTGTGTGTAAGAATGA 3547
QY 3669 TGATGGAATCTACAGACCCAGAGCTATCAGTATTCAATCAGCCAGCAAGAGTACTGG 3728
Db 3548 TGATGGAATCTACAGACCCAGAGCTATCAGTATTCAATCAGCCAGCAAGAGTACTGG 3507
QY 3729 TGCCAGTCTTCTTGTGTAGTGGCGCTCTGAAATCCTCTTCTGGATACCTTCCCAAGTC 3788
Db 3608 TGCCAGTCTTCTTGTGTAGTGGCGCTCTGAAATCCTCTTCTGGATACCTTCCCAAGTC 3667
QY 3789 CAGTATTGTGAAGATGGTGTATGTTCCAGATTACTGCAGAGCAATGATTCCTTGAG 3848
Db 3668 CAGTATTGTGAAGATGGTGTATGTTCCAGATTACTGCAGAGCAATGATTCCTTGAG 3727
QY 3849 GCAGGCACTGGAGAGATGAAGGACTTCCACATCAGCTGGGAGGCGGACGCGAGGA 3908
Db 3728 GCAGGCACTGGAGAGATGAAGGACTTCCACATCAGCTGGGAGGCGGACGCGAGGA 3787
QY 3909 ACCCAGAGCACATCCACATCCAGTGGGTGATGATGACAAAGAACCTTAGCAAGGGTGT 3968
Db 3788 ACCCAGAGCACATCCACATCCAGTGGGTGATGATGACAAAGAACCTTAGCAAGGGTGT 3847
QY 3969 CGTAAGTCCATAGATGGAGTCCAGTGGAGCTATTAACAAATGTTGAAGATATTCATGG 4028
Db 3848 CGTAAGTCCATAGATGGAGTCCAGTGGAGCTATTAACAAATGTTGAAGATATTCATGG 3907
QY 4029 ATCAGAATATAAGCAAAATGAAAGTAATCAGATGGACAGAGGTGTTTTCTTAGAAA 4088
Db 3908 ATCAGAATATAAGCAAAATGAAAGTAATCAGATGGACAGAGGTGTTTTCTTAGAAA 3967
QY 4089 CGATGACAGACAAATTCCTCAGTATCTCGAGATCAGATGATGATGATGATGATGATG 4148
Db 3968 CGATGACAGACAAATTCCTCAGTATCTCGAGATCAGATGATGATGATGATGATGATG 4027
QY 4149 TGCCAAAGCTTTTGGCTTGTCTCTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 4208
Db 4028 TGCCAAAGCTTTTGGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4087
QY 4209 GACCAACTGGGACTAGCTGTGACACTTGAATCAGATCAGATGATGATGATGATGATGATG 4268
Db 4088 GACCAACTGGGACTAGCTGTGACACTTGAATCAGATCAGATGATGATGATGATGATGATG 4147
QY 4269 CAATGGCCAGCCCTTCCCTCGCAGTACATCAATGATCTGATAGCCCTTGGTGGCCGT 4328
Db 4148 CAATGGCCAGCCCTTCCCTCGCAGTACATCAATGATCTGATAGCCCTTGGTGGCCGT 4207
QY 4329 GATCCATGGAGGGCTGCCAGCTTAGTGGGCCCCCGTGTGATGAACTCATCTTTTA 4388
Db 4208 GATCCATGGAGGGCTGCCAGCTTAGTGGGCCCCCGTGTGATGAACTCATCTTTTA 4267
QY 4389 TATTCGAAAACATCGTATAAAGAGAGACATCTATTTTTTCTGTCAGACTTCTTG 4448
Db 4268 TATTCGAAAACATCGTATAAAGAGAGACATCTATTTTTTCTGTCAGACTTCTTG 4327
QY 4449 CAACAGAGTATACCCCAATCAATTTGACATTTTAAACTGGAAGATTAAGCTTTTAA 4508
Db 4328 CAACAGAGTATACCCCAATCAATTTGACATTTTAAACTGGAAGATTAAGCTTTTAA 4387

QY 4509 CACTATTATGGGTGGGGAATAGGTTGGAGTAGGGGTTTGGGAGACGGTGGGAAGG 4568
Db 4388 CACTATTATGGGTGGGGAATAGGTTGGAGTAGGGGTTTGGGAGACGGTGGGAAGG 4447
QY 4569 GTGGTTGGGGGACCGATGTTCCATAATTTCTAAGTCTTCTATGATGATTCACCAAGAAG 4628
Db 4448 GTGGTTGGGGGACCGATGTTCCATAATTTCTAAGTCTTCTATGATGATTCACCAAGAAG 4507
QY 4629 APTCTGGGAGCTTCTGTTCTCTGCAACACAGTATGCTATCTTGGCAGTATTCCTCTCT 4688
Db 4508 ATCTGGGAGCTTCTGTTCTCTGCAACACAGTATGCTATCTTGGCAGTATTCCTCTCT 4567
QY 4689 GTTACTGTTTACAGCAAGATTCCTGCTCTCTCTCAAGATTTACTTATGCTATGCTCA 4748
Db 4568 GTTACTGTTTACAGCAAGATTCCTGCTCTCTCTCTCAAGATTTACTTATGCTATGCTCA 4627
QY 4749 GAAATGCTCAATGGGTACCAACCATCACCAAGGTGGATGGAGGGGAGGGAATA 4808
Db 4628 GAAATGCTCAATGGGTACCAACCATCACCAAGGTGGATGGAGGGGAGGGAATA 4687
QY 4809 AAATATAAGCATCA 4823
Db 4688 AAATATAAGCATCA 4702

RESULT 3
AF130419 4565 bp mRNA linear PRI 23-MAY-1999
LOCUS Homo sapiens serine protease-like protein isoform (NSP) mRNA,
DEFINITION alternatively spliced, complete cds.
ACCESSION AF130419
VERSION AF130419.1 GI:4883895
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 4565)
AUTHORS Meckelein,B., Marshall,D.C., Conn,K.J., Pietropaolo,M., Van
Nostrand,W. and Abraham,C.R.
TITLE Identification of a novel serine protease-like molecule in human
Brain
JOURNAL Brain Res. 55 (2), 181-197 (1998)
MEDLINE 98250799
PUBMED 9582421
REFERENCE 2 (bases 1 to 4565)
AUTHORS Meckelein,B., Marshall,D.C.L., Conn,K.J., Pietropaolo,M., Van
Nostrand,W. and Abraham,C.R.
TITLE Direct Submission
JOURNAL Submitted (22-FEB-1999) Biochemistry, Boston University Medical
School, 715 Albany St., Bldg. K621, Boston, MA 02118, USA
FEATURES
Location/Qualifiers
1..4565
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
/tissue_type="brain"
/dev_stage="fetal"
1..4565
/gene="NSP"
173..3805
/gene="NSP"
/note="alternatively spliced; isolated with antibody
against Homo sapiens cathepsin G; involved in Alzheimer's
disease"
/codon_start=1
/product="serine protease-like protein isoform"
/protein_id="AA031694.1"
/db_xref="GI:4883895"
/translation="MENYFQAEYANLDKVLDFEONEEDTVSTLLDTKWKILDPSPS
HRLSNPTLASVNESAVNESQFLKVFSLASAPLTTEEDHCANGQCNLPBIAT
MWIDENAVAEQDLIKRNSWDDQCSAVEVEGKCKGNLACLPEKIVLVVAVMHNCCKR

TLONLDQDNYNYSOLMDAFSCSLDNRQDQFSPSINESTKDMSEKQMDPLNR
 PRTEGRSVNHLCPITSSDLSVCSQSPQKDDGSGRDPMSAITSIVDSVISQGD
 GCPAVKQENYIPDSDLTKISSPRTDLGSPNSFSHMSSEGLMKKEPAEESTESUR
 SGLPILLKPDNPNGSGRNDRCSDCLVPNEVRADENEGYEHEETLGTTEFLNMTSH
 FSESDMTNWKLTNLMNDSOVNBEKEKFLQISOPBDTNGDSGQCVGLADGLDLK
 GTCISESECDFTVIDTPAANYLSNGCDSYGMODPGVSFVPKTLPSKEDSVTEKEI
 RESKSECVSYNEORGNATEGSGLLNSTGDLAKKNYLHNFCQVPSVLGQSPKV
 ASLPSVFPGARQKPSNKLQIPKPLSDHLQNDPPANSNNITKNNDILGKAKUG
 ENSATNCSPLSGNLSNDVTNGHLESYEASISTRPCIALAPDPSDNDLRAGQFISA
 RKPFTLGEVAPVWVDSQAFNCKCEARFTTKRHRHCRACGVKFCASCCSKCKLL
 YMDREARVCVICHSLVNAQPREQVFPADGILPNEVADAALKTNGTSSAGTL
 AVSHDEKPVVTSPLPAETDCLFSGSITQVSPVGMNLIPEGLPPIILISGVGK
 DYAVEEKPSQJLVMOLEDGDPDLVFLVNLALSMVKIYVNVNRCWCFITTKGMHVA
 QGSEBIVILLQCLPEKCLPXDIFNHFVQLYDALAGNVSVNLGSHSPSQSFLSKSEHG
 GLVYVTSYQSLQDLVLPFLYFLLGILQKWEIPEWKFVPIRLMLRIGAEIRYPCPL
 FSVPRKPLFGETGHTIMNLADPNFYQYTLFVVOGLVDMVEVNRKTSIKIPSNRYNEM
 MKAMNKSHEVLGAGACNEKADSHLVQVQNDGNYQTOAISIHNPQKVTGASFFVF
 SGALKSSSYLAKSSIVBDGVVWVQITAEINMDSLRQALRDEGLHHHLWEGRGSTPGAH
 PHPVGG"

misc_feature 1475..1477

/gene="NSP"

/note="active site"

ORIGIN

Query Match 89.1%; Score 4309.2; DB 9; Length 4565;

Best Local Similarity 96.1%; Pred. No. 0;

Matches 4521; Conservative 0; Mismatches 3; Indels 180; Gaps 3;

QY	1	GCATCTGAATCAGCAGCACTGGCTGGTGGTGCAGCAGACATCATGAGTAAGCACCGAGA	60
DB	42	GCATCTGAATCAGCAGCACTGGCTGGTGGTGCAGCAGACATCATGAGTAAGCACCGAGA	101
QY	61	AGTCTGTTCTTATCACCTGTGAAGGGGAAAAGGTTTAAACAAGTCTCTTAAGTGGTG	120
DB	102	AGTCTGTTCTTATCACCTGTGAAGGGGAAAAGGTTTAAACAAGTCTCTTAAGTGGTG	161
QY	121	TTTCTCCAGGATGAGAAATTAATTCAGAGCAAGCTTCAACCTGGGCAAGGTGTTA	180
DB	162	TTTCTCCAGGATGAGAAATTAATTCAGAGCAAGCTTCAACCTGGGCAAGGTGTTA	220
QY	181	GATGAATTTGAACAAACGAGATGAACAGTTCTTCTACTTCTTATTTGGATACAAAGTGG	240
DB	221	GATGAATTTGAACAAACGAGATGAACAGTTCTTCTACTTCTTATTTGGATACAAAGTGG	280
QY	241	AATGAAGTTCTAGATCCCTCTTCCCGGCTGTCATTTAAACCTACTTTGGCAGGTGG	300
DB	281	AATGAAGTTCTAGATCCCTCTTCCCGGCTGTCATTTAAACCTACTTTGGCAGGTGG	340
QY	301	AATGAATCTGAGTTCTTAATGAGTCACAACTCAACCTGAAAGTCTTCTCCCTGGCTCAT	360
DB	341	AATGAATCTGAGTTCTTAATGAGTCACAACTCAACCTGAAAGTCTTCTCCCTGGCTCAT	400
QY	361	TCAGTCTCCCTGACACAGAGAGAGGATCACTGTGCTAATGACAGGACTGTGAATCTTA	420
DB	401	TCAGTCTCCCTGACACAGAGAGAGGATCACTGTGCTAATGACAGGACTGTGAATCTTA	460
QY	421	AATCCAGAGATTGCCAATGTGATTTGATGAATAATGCTGTTCAGAGAAGACAGTTAATT	480
DB	461	AATCCAGAGATTGCCAATGTGATTTGATGAATAATGCTGTTCAGAGAAGACAGTTAATT	520
QY	481	AAGAGAACTATAGTTGGATGATCAATGACAGTCTGTGTTGAAGTGGGAGAGAGAATGT	540
DB	521	AAGAGAACTATAGTTGGATGATCAATGACAGTCTGTGTTGAAGTGGGAGAGAGAATGT	580
QY	541	GGAAACCTGGCTGTGTCAGATGAGAGAATGTTCTTGTGTGAGCCGTCATGCAATAAC	600
DB	581	GGAAACCTGGCTGTGTCAGATGAGAGAATGTTCTTGTGTGAGCCGTCATGCAATAAC	640
QY	601	TGTGATAAAGGACATTACAAAACGATTTCAGGATTTGTAATTAATATAGTCAATCC	660
DB	641	TGTGATAAAGGACATTACAAAACGATTTCAGGATTTGTAATTAATATAGTCAATCC	700
QY	661	CTTATGAGATGCTTTTAGCTGTTCACTGGATAATGAAAACAGACAAACTGATCAATTTAGT	720

DB	701	CTTATGAGATGCTTTTAGCTGTTTCACTGGATAATGAAAACAGACAACTGATCAATTTAGT	760
QY	721	TTTAGTATAAATGAGTCCACTGAAAAGATATGAATTCAGAGAAAACAAATGGATCCATTG	780
DB	761	TTTAGTATAAATGAGTCCACTGAAAAGATATGAATTCAGAGAAAACAAATGGATCCATTG	820
QY	781	AATAGACCGAAAACAGAGGGGAGATCTGTTAAACCATCTGTGTCCTACTTCATCTGATAGT	840
DB	821	AATAGACCGAAAACAGAGGGGAGATCTGTTAAACCATCTGTGTCCTACTTCATCTGATAGT	880
QY	841	CTAGCCAGTGTCTGTTTCCCTCTTCAAAATTAAGGATGACGGAAGTATAGGTAGAGACCCC	900
DB	881	CTAGCCAGTGTCTGTTTCCCTCTTCAAAATTAAGGATGACGGAAGTATAGGTAGAGACCCC	940
QY	901	TCCATGCTGCGATTAACAAGTTTAAACGGTTGATTCAGTAATCTCATCCCAGGAAACAGAT	960
DB	941	TCCATGCTGCGATTAACAAGTTTAAACGGTTGATTCAGTAATCTCATCCCAGGAAACAGAT	1000
QY	961	GGATGCTCTGCTGTTTAAAAAGCAAGAACTATATACAGATGAGGACCTCACTGGCAAA	1020
DB	1001	GGATGCTCTGCTGTTTAAAAAGCAAGAACTATATACAGATGAGGACCTCACTGGCAAA	1060
QY	1021	ATCAGCTCTCTTAGGACAGATCTAGGAGTCCAAATTCCTTTCCCATGAGTGAGGGG	1080
DB	1061	ATCAGCTCTCTTAGGACAGATCTAGGAGTCCAAATTCCTTTCCCATGAGTGAGGGG	1120
QY	1081	ATTTCATGAAAAAGAGCCAGCAGGAGAGCACCCTGAAGATCCCTCCGGTCTGGT	1140
DB	1121	ATTTCATGAAAAAGAGCCAGCAGGAGAGCACCCTGAAGATCCCTCCGGTCTGGT	1180
QY	1141	TTACCTTCTCTCTCAAAACACAGATGCTCTAATGGGTCTGGAAGGAATAATGACTGTCAA	1200
DB	1181	TTACCTTCTCTCTCAAAACACAGATGCTCTAATGGGTCTGGAAGGAATAATGACTGTCAA	1240
QY	1201	CGGTGTTTCAGATGCTTGTGCTTAATGAAGTTAGGCTGTATGAATAATGAAGTTATGAA	1260
DB	1241	CGGTGTTTCAGATGCTTGTGCTTAATGAAGTTAGGCTGTATGAATAATGAAGTTATGAA	1300
QY	1261	CATGAAGAACTCTTGGCACTACAGAAATTCCTTAATATGACAGAGACTTCTCTGAATCT	1320
DB	1301	CATGAAGAACTCTTGGCACTACAGAAATTCCTTAATATGACAGAGACTTCTCTGAATCT	1360
QY	1321	CAGGACATGCTAATTTGGAAGTTGCTAAACTTAATAGATGATGATGATGATGATGATGATG	1380
DB	1361	CAGGACATGCTAATTTGGAAGTTGCTAAACTTAATAGATGATGATGATGATGATGATGATG	1420
QY	1381	GAGAAAAGGAAAAGTTTCTACAGATTTAGTCAGCTGAGGACACTAATGGTGTAGTGGGA	1440
DB	1421	GAGAAAAGGAAAAGTTTCTACAGATTTAGTCAGCTGAGGACACTAATGGTGTAGTGGGA	1480
QY	1441	GGACAGTGTGTTGGATTGGCAGATCGAGTCTAGATTTTAAAGGAACTTGCATTAAT	1500
DB	1481	GGACAGTGTGTTGGATTGGCAGATCGAGTCTAGATTTTAAAGGAACTTGCATTAAT	1540
QY	1501	AGTGAAGATGATTTCTCCACTTATAGACACACAGCAGCAGCAAAATTTCTATCAAT	1560
DB	1541	AGTGAAGATGATTTCTCCACTTATAGACACACAGCAGCAGCAAAATTTCTATCAAT	1600
QY	1561	GTTTGTGATTCCTATGGAATGCAAGACCCAGGTGTTCTTTTGTTCCTAAAGACTTTACCC	1620
DB	1601	GTTTGTGATTCCTATGGAATGCAAGACCCAGGTGTTCTTTTGTTCCTAAAGACTTTACCC	1660
QY	1621	TCCAAAGAAGATTCAGTAAACAGAGAAAAGAAATAGAGAAAGCAAGTCAAGATGCTAC	1680
DB	1661	TCCAAAGAAGATTCAGTAAACAGAGAAAAGAAATAGAGAAAGCAAGTCAAGATGCTAC	1720
QY	1681	TCAAAATATTTATGAACAGAGAGAAATCAGGCCCACAGAGAGGAGTGGACTACTTTTAAAC	1740
DB	1721	TCAAAATATTTATGAACAGAGAGAAATCAGGCCCACAGAGAGGAGTGGACTACTTTTAAAC	1780
QY	1741	AGCAGTGGTACCTAATGAAGAAAATTTATTATCAATAATTTCTGTAGTCAAGTTCATCA	1800

QY 1581 GCAAGACCAGGTTCTTTTGTTCCTCAAGACTTTTACCTCCAAAGAGATTCAGTAAC 1640
 Db 61467 GCAAGACCAGGTTCTTTTGTTCCTCAAGACTTTTACCTCCAAAGAGATTCAGTAAC 61526
 QY 1641 AGAAGAAAAGAAATAGAGAAAGCAAGTCAAGTCTACTCAATATTTATGAACAGAG 1700
 Db 61527 AGAAGAAAAGAAATAGAGAAAGCAAGTCAAGTCTACTCAATATTTATGAACAGAG 61586
 QY 1701 AGAATAGGCGCACAGAGGAGTGGACTACTTTTAAACACACTGTCACCTTAATGAA 1760
 Db 61587 AGAATAGGCGCACAGAGGAGTGGACTACTTTTAAACACACTGTCACCTTAATGAA 61646
 QY 1761 GAAAATATTTACATAAATTTCTGTAGTCAAGTTCATCAGTCTTGGCAATCTTCCCC 1820
 Db 61647 GAAAATATTTACATAAATTTCTGTAGTCAAGTTCATCAGTCTTGGCAATCTTCCCC 61706
 QY 1821 CAAGTAGTAGCAAGCTCCCATCTATCAGTCTTCTTTGGTGTGCAAGACCAAGCA 1880
 Db 61707 CAAGTAGTAGCAAGCTCCCATCTATCAGTCTTCTTTGGTGTGCAAGACCAAGCA 61766
 QY 1881 ACCTTCTAATCTAACTTCAATTCCTCAAGGCAATATCAGACCAATTTACAAAATGACTT 1940
 Db 61767 ACCTTCTAATCTAATCTCAATTCCTCAAGGCAATATCAGACCAATTTACAAAATGACTT 61826
 QY 1941 TCTGCAACAGTGAATAATACTAATAAATAAATAAATATCTTCTTGGAAAGCAAAAT 2000
 Db 61827 TCTGCAACAGTGAATAATACTAATAAATAAATAAATATCTTCTTGGAAAGCAAAAT 61886
 QY 2001 AGGGGAAACTCAGCAACCAATGTAGTGCAGTCCATCTTTGGAAACATCTTATGTGCA 2060
 Db 61887 AGGGGAAACTCAGCAACCAATGTAGTGCAGTCCATCTTTGGAAACATCTTATGTGCA 61946
 QY 2061 TACAAATGGGAAACATTTAGAAAGTTATGAGCTGAGATCTCCATAGACCATGCTTGC 2120
 Db 61947 TACAAATGGGAAACATTTAGAAAGTTATGAGCTGAGATCTCCATAGACCATGCTTGC 62006
 QY 2121 ATTACTCCAGATAGCCAGATATGATCTCAGAGCTGGTCAAGTTTGAATTTCTGCGAG 2180
 Db 62007 ATTACTCCAGATAGCCAGATATGATCTCAGAGCTGGTCAAGTTTGAATTTCTGCGAG 62066
 QY 2181 AAAGCATTACACAGCTGGGTGAGTGGCTCCAGTATGGGTACCGGATCTCAGGCTCC 2240
 Db 62067 AAAGCATTACACAGCTGGGTGAGTGGCTCCAGTATGGGTACCGGATCTCAGGCTCC 62126
 QY 2241 AAATTCATGATAGTGAAGCCAGTTTACATTCACAAAAGGAGGATCATCTCAGAGC 2300
 Db 62127 AAATTCATGATAGTGAAGCCAGTTTACATTCACAAAAGGAGGATCATCTCAGAGC 62186
 QY 2301 ATGTGGGAAGGT 2312
 Db 62187 ATGTGGGAAGGT 62198
 RESULT 6
 AC026909
 LOCUS 150129 bp DNA linear HTG 19-APR-2000
 DEFINITION Homo sapiens chromosome 1 clone RP11-174F10 map 1, WORKING DRAFT
 SEQUENCE, 27 unordered pieces.
 AC026909
 VERSION AC026909.2 GI:7596849
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 150129)
 Biren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens chromosome 1, clone RP11-174F10
 JOURNAL Unpublished
 2 (bases 1 to 150129)
 Biren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,

Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeAureliano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
 Galand, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Klein, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Levine, R., Lie, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrum, J., Menues, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tessier, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 19, 2000 this sequence version replaced gi:7328772.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIER
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L7273
 Center clone name: 174_F_10
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 134873 bases at least Q40
 Consensus quality: 142261 bases at least Q30
 Consensus quality: 145325 bases at least Q20
 Insert size: 155000; agarose-fp
 Insert size: 147529; sum-of-contigs
 Quality coverage: 3.7 in Q20 bases; agarose-fp
 Quality coverage: 3.9 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 27 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1572: contig of 1572 bp in length
 * 1573 1672: gap of 100 bp
 * 1673 2741: contig of 1069 bp in length
 * 2742 2841: gap of 100 bp
 * 2842 4870: contig of 2029 bp in length
 * 4871 4971: gap of 100 bp
 * 4971 6056: contig of 1086 bp in length
 * 6057 6157: gap of 100 bp
 * 6157 7419: contig of 1263 bp in length
 * 7420 7519: gap of 100 bp
 * 7520 9305: contig of 1786 bp in length
 * 9306 9406: gap of 100 bp
 * 9406 11199: contig of 1794 bp in length
 * 11200 11299: gap of 100 bp
 * 11300 13054: contig of 1755 bp in length
 * 13055 13154: gap of 100 bp
 * 13155 15817: contig of 2663 bp in length
 * 15818 15917: gap of 100 bp
 * 15918 18265: contig of 2348 bp in length


```

Db 70675 GAGATCTGTTAACCATCTGTGCTCTACTCTCACTGATGCTAGCCAGTGTCTGTTCCCC 70734
QY 861 TTCCAAATTAAGGATGACGGAAGTATAGGTAGAGACCCCTCCATCTGTCTGCGATTACAAG 920
Db 70735 TTCCAAATTAAGGATGACGGAAGTATAGGTAGAGACCCCTCCATCTGTCTGCGATTACAAG 70734
QY 921 TTTAACGGTGTGATTCAGTAATCTCATCCAGGACACAGATGATGCTCTGCTGTTAAAAA 980
Db 70795 TTTAACGGTGTGATTCAGTAATCTCATCCAGGACACAGATGATGCTCTGCTGTTAAAAA 70854
QY 981 GCAAGAGAACTATATACCAAGATGAGGACCTCACTGCGAAAATCAGCTCTCTCTAGGACAGA 1040
Db 70855 GCAAGAGAACTATATACCAAGATGAGGACCTCACTGCGAAAATCAGCTCTCTCTAGGACAGA 70914
QY 1041 TCTAGGGAGTCCAAATTCCTTTTCCCAATGATGAGGGGATTTGATGAAAAAGAGCC 1100
Db 70915 TCTAGGGAGTCCAAATTCCTTTTCCCAATGATGAGGGGATTTGATGAAAAAGAGCC 70974
QY 1101 AGCAGAGGAGAGACACCACTGAAGAATCCCTCGGCTCTGCTTTACCTTTGCTTCTCAAAACC 1160
Db 70975 AGCAGAGGAGAGACACCACTGAAGAATCCCTCGGCTCTGCTTTACCTTTGCTTCTCAAAACC 71034
QY 1161 AGACATGCTTAATGGGTCTGGAAGGAATAATGACTGTGAACGGTGTTCAGATTGCTTGT 1220
Db 71035 AGACATGCTTAATGGGTCTGGAAGGAATAATGACTGTGAACGGTGTTCAGATTGCTTGT 71094
QY 1221 GCCTAATGAAGTTAGGGCTGATGAAATGAAGTTATGAACATGAAGAACTCTTGGCAC 1280
Db 71095 GCCTAATGAAGTTAGGGCTGATGAAATGAAGTTATGAACATGAAGAACTCTTGGCAC 71154
QY 1281 TACAGAATTCCTTAATATACAGAGCAATTTCTGAAATCTCAGGACATGACTAATTTGAA 1340
Db 71155 TACAGAATTCCTTAATATACAGAGCAATTTCTGAAATCTCAGGACATGACTAATTTGAA 71214
QY 1341 GTTGACTAACTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1400
Db 71215 GTTGACTAACTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 71274
QY 1401 ACAGATTAGTCAGGCTGAGGACACTTAATGTTGATGAGGACAGTGTGTTGATTTGGC 1460
Db 71275 ACAGATTAGTCAGGCTGAGGACACTTAATGTTGATGAGGACAGTGTGTTGATTTGGC 71334
QY 1461 AGATCAGGCTTAGATTAAAGGAACCTTGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1520
Db 71335 AGATCAGGCTTAGATTAAAGGAACCTTGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 71394
QY 1521 CACTGTTATAGACACACAGCAGCAAAATTAATCTATCTAATGTTGATGATTCCTATGGAAT 1580
Db 71395 CACTGTTATAGACACACAGCAGCAAAATTAATCTAATGTTGATGATTCCTATGGAAT 71454
QY 1581 GCAGAACCCAGGTGTTCTTTTGTCCAAAGACTTTTACCCTCCAAAGAGATTTCAGTAAC 1640
Db 71455 GCAGAACCCAGGTGTTCTTTTGTCCAAAGACTTTTACCCTCCAAAGAGATTTCAGTAAC 71514
QY 1641 AGAAGAAAAGAAATAGAGGAAGCAAGTCAAGATGCTACTCAAAATTTATGAACAGAG 1700
Db 71515 AGAAGAAAAGAAATAGAGGAAGCAAGTCAAGATGCTACTCAAAATTTATGAACAGAG 71574
QY 1701 AGGAAATGAGCCACAGAGGGAGTGGACTACTTTTAAACAGCAGCTGGTGACCTAATGAA 1760
Db 71575 AGGAAATGAGCCACAGAGGGAGTGGACTACTTTTAAACAGCAGCTGGTGACCTAATGAA 71634
QY 1761 GAAAAATTTATACATAATTTCTAGTCAAGTTCATCAGTGTGGGCAATCTTCCCC 1820
Db 71635 GAAAAATTTATACATAATTTCTAGTCAAGTTCATCAGTGTGGGCAATCTTCCCC 71694
QY 1821 CAAGGTAGTAGCAAGCCTGCGCATCTATCAGTGTCTCTTTTGGTGGTGAAGACCCCAAGCA 1880
Db 71695 CAAGGTAGTAGCAAGCCTGCGCATCTATCAGTGTCTCTTTTGGTGGTGAAGACCCCAAGCA 71754
QY 1881 ACCTTCTAACTTAACTTCAATTCCAAAGCCATATACAGACATTTCAAAATGACTTT 1940

```

```

Db 71755 ACCTTCTAACTTAACTTCAAAATTCAAAAGCCATTATCAGACCATTTACAAAATGACTT 71814
QY 1941 TCCTGCAAAACAGTGGAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2000
Db 71815 TCCTGCAAAACAGTGGAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 71874
QY 2001 AGGGAAAACTCAGCAACCAATGATGTCAGTGCATCTTTGGGAAAAATCTCTAATGTCGA 2060
Db 71875 AGGGAAAACTCAGCAACCAATGATGTCAGTGCATCTTTGGGAAAAATCTCTAATGTCGA 71934
QY 2061 TACAAATGGGAAAACTTTAGAAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCTTCC 2120
Db 71935 TACAAATGGGAAAACTTTAGAAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCTTCC 71994
QY 2121 ATTAGTCTCAGATAGCCCAAGATAATGATCTCAGAGCTGCTCAGTTTGGAAATTTCTGCCAG 2180
Db 71995 ATTAGTCTCAGATAGCCCAAGATAATGATCTCAGAGCTGCTCAGTTTGGAAATTTCTGCCAG 72054
QY 2181 AAAGCCATTACCAACCTCGGTGAGTGGCTCCAGTATGGTACCGGATTTCTCAGGCTCC 2240
Db 72055 AAAGCCATTACCAACCTCGGTGAGTGGCTCCAGTATGGTACCGGATTTCTCAGGCTCC 72114
QY 2241 AAATTCATGAATGTGAAGCCAGGTTTACATTCACAAAAGAGGAGCATCACTGCAGAGC 2300
Db 72115 AAATTCATGAATGTGAAGCCAGGTTTACATTCACAAAAGAGGAGCATCACTGCAGAGC 72174
QY 2301 ATCTGGGAAGGT 2312
Db 72175 ATCTGGGAAGGT 72186

RESULT 7
AX809313
LOCUS AX809313
DEFINITION Sequence 10 from Patent WO03045990.
ACCESSION AX809313
VERSION AX809313.1 GI:38523628
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Legrain, P., Gauthier, J.M., Colland, F. and Jacq, X.
TITLE Protein-protein interactions involving transforming growth factor
beta signalling
JOURNAL Patent: WO 03045990-A 10 05-JUN-2003;
Hybridomics (PR)
FEATURES
source
location/Qualifiers
1..1977
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 40.9%; Score 1977; DB 6; Length 1977;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2434 AGTGCTCAAGCCAGAGCCCTAACCTCAACAATCTCTGTAATCTGTTCTACTATCCCT 2493
Db 1 AGTGCTCAAGCCAGAGCCCTAACCTCAACAATCTCTGTAATCTGTTCTACTATCCCT 60
QY 2494 CCTTGCAGCAAGCTCAGGCTCAGGCTCTGAGCTCTCCACTCCCACTGTGATGTA 2553
Db 61 CCTTGCAGCAAGCTCAGGCTCAGGCTCTGAGCTCTCCACTCCCACTGTGATGTA 120
QY 2554 CCTGTTGGAGTTTAAAGCACCTTGGAGCAAGTGTCTAGCCAGAGGAGGCGGA 2613
Db 121 CCTGTTGGAGTTTAAAGCACCTTGGAGCAAGTGTCTAGCCAGAGGAGGCGGA 180
QY 2614 GTTTGGTTGCTGATGGGATCTTCCCAATGGAGAGTGTCTGATGAGCAAAATTAACA 2673

```

Db	181	GTCTGGTTGCTGATGGATCTTTGCCAATGGAGAGTTGCTGATGACGCCAAATTAACA	240
Qy	2674	ATGAATGAATCTTCTCTGAGAAACCTGGCTGTGTACACAGCCAGTCAAGCCAGTA	2733
Db	241	ATGAATGAATCTTCTCTGAGAAACCTGGCTGTGTACACAGCCAGTCAAGCCAGTA	300
Qy	2734	ACTACAGTCTCTTACACAGAGAGCGATATTTGTCTATTCTCTGGAGTATAAATCAG	2793
Db	301	ACTACAGTCTCTTACACAGAGAGCGATATTTGTCTATTCTCTGGAGTATAAATCAG	360
Qy	2794	GTTCGAAGTCTCTGTGGAAGTGAATGAATCTTATTCCTGAAGTGGCTTCTCCCAT	2853
Db	361	GTTCGAAGTCTCTGTGGAAGTGAATGAATCTTATTCCTGAAGTGGCTTCTCCCAT	420
Qy	2854	CTCATCTCCACTGGTGAAGGAGACTATGCTGTGGAAGAGAAACCATCACAGATTCA	2913
Db	421	CTCATCTCCACTGGTGAAGGAGACTATGCTGTGGAAGAGAAACCATCACAGATTCA	480
Qy	2914	GTATGACAGAGTTGGAGGATGTGGCCCTGACCCACTTGTATTTGTTTTAAATGCAAT	2973
Db	481	GTATGACAGAGTTGGAGGATGTGGCCCTGACCCACTTGTATTTGTTTTAAATGCAAT	540
Qy	2974	TTGTTGTCATGTTTAAATTTGAATTTATGTGAACAGGAGTGTGGTGTTCACAACC	3033
Db	541	TTGTTGTCATGTTTAAATTTGAATTTATGTGAACAGGAGTGTGGTGTTCACAACC	600
Qy	3034	AAGGGAATGCATGCAGTGGGTGCTGAGATAGTCACTTCTTACAGTGTTTACCGAT	3093
Db	601	AAGGGAATGCATGCAGTGGGTGCTGAGATAGTCACTTCTTACAGTGTTTACCGAT	660
Qy	3094	GAAGAAGTGTTCGCAAGGATATCTTTAAATCACTTTGTGAGCTTTACGGATGCTCTG	3153
Db	661	GAAGAAGTGTTCGCAAGGATATCTTTAAATCACTTTGTGAGCTTTACGGATGCTCTG	720
Qy	3154	GCAGGAATGTGTGAGCACTGGGACATCTCTTCTTCACTCAAGATTTCTTGGCAGT	3213
Db	721	GCAGGAATGTGTGAGCACTGGGACATCTCTTCTTCACTCAAGATTTCTTGGCAGT	780
Qy	3214	AAAGAACATGTTGATTTCTTATATGTGACATCTTACCTACAGTCACTGCAAGACTAGTA	3273
Db	781	AAAGAACATGTTGATTTCTTATATGTGACATCTTACCTACAGTCACTGCAAGACTAGTA	840
Qy	3274	CTCCCAACCCACCTTACTTGTGGGATCTTATCCAGAAATGGGAACCTCCTTGGGCT	3333
Db	841	CTCCCAACCCACCTTACTTGTGGGATCTTATCCAGAAATGGGAACCTCCTTGGGCT	900
Qy	3334	AAAGTATTTCTATCCGCTGATGTGAGACTTGGAGCTGAATATCGACTTTATCCATGC	3393
Db	901	AAAGTATTTCTATCCGCTGATGTGAGACTTGGAGCTGAATATCGACTTTATCCATGC	960
Qy	3394	CCACTATTCAGTGCAGATTTGGAGGCCATTTGTTGGAGAGCGGGGCATACCATCATG	3453
Db	961	CCACTATTCAGTGCAGATTTGGAGGCCATTTGTTGGAGAGCGGGGCATACCATCATG	1020
Qy	3454	AAATCTCTTGAGACTTCAGAAATACAGATATACCTTCCAGTAGTTCAGAGTTGGTG	3513
Db	1021	AAATCTCTTGAGACTTCAGAAATACAGATATACCTTCCAGTAGTTCAGAGTTGGTG	1080
Qy	3514	GTTCATGGAAGTTCCGAAACTAGCATCAAAATCCCAAGCAACAGATCAATGAGATG	3573
Db	1081	GTTCATGGAAGTTCCGAAACTAGCATCAAAATCCCAAGCAACAGATCAATGAGATG	1140
Qy	3574	ATGAAGCATGAACAAAGTCCATGAGCATGTCTCGCAGGAGTGTCTGCTTCAATGAA	3633
Db	1141	ATGAAGCATGAACAAAGTCCATGAGCATGTCTCGCAGGAGTGTCTGCTTCAATGAA	1200
Qy	3634	AAGGAGACTCTCATCTTGTGTGTACAGATGATGGAATATGGAATATCAGACCCAGCT	3693
Db	1201	AAGGAGACTCTCATCTTGTGTGTACAGATGATGGAATATGGAATATCAGACCCAGCT	1260
Qy	3694	ATCAGTATTCAAATCAGCCCAAGAAAGTACTGTGTCAGTTCTTGTGTTGAGTGGC	3753
Db	1261	ATCAGTATTCAAATCAGCCCAAGAAAGTACTGTGTCAGTTCTTGTGTTGAGTGGC	1320

Qy	3754	GCCTGAAATCTCTTCTGATACCTTCCCAAGTCCAGTATTGGAAGATGGTGTATG	3813
Db	1321	GCCTGAAATCTCTTCTGATACCTTCCCAAGTCCAGTATTGGAAGATGGTGTATG	1380
Qy	3814	GTCCAGATCTGACAGAAACATGATTCCTTGGAGGAGGACCTGCGAGATGAAGAC	3873
Db	1381	GTCCAGATCTGACAGAAACATGATTCCTTGGAGGAGGACCTGCGAGATGAAGAC	1440
Qy	3874	TTCCACCATCACCTGTGGGAAGCGGACCGGAGGAACCCAGGAGCACATCCACATCCAG	3933
Db	1441	TTCCACCATCACCTGTGGGAAGCGGACCGGAGGAACCCAGGAGCACATCCACATCCAG	1500
Qy	3934	TGGGTGGATGATGACAGAAACGTTAGCAAGGGTGTGTAAGTCTTATAGATGGGAAGTCC	3993
Db	1501	TGGGTGGATGATGACAGAAACGTTAGCAAGGGTGTGTAAGTCTTATAGATGGGAAGTCC	1560
Qy	3994	ATGGAGACTATACAAATGTGAAGATATTCATGGATCAGATATATAAGCAAAATGGAATA	4053
Db	1561	ATGGAGACTATACAAATGTGAAGATATTCATGGATCAGATATATAAGCAAAATGGAATA	1620
Qy	4054	GTAAATCAGATGACAGAGTGTTCCTAGAAAAAGATGACCAACATTCGCTCAGT	4113
Db	1621	GTAAATCAGATGACAGAGTGTTCCTAGAAAAAGATGACCAACATTCGCTCAGT	1680
Qy	4114	GATCTGAGATCAGATAGATTTGACTGAGCATGTGCAAGATTTTGCCTTCTCTC	4173
Db	1681	GATCTGAGATCAGATAGATTTGACTGAGCATGTGCAAGATTTTGCCTTCTCTC	1740
Qy	4174	TGTCTCTCACTCAAACTTCTGAAGGAAGATGAAATGACCAAACTGGGACTACGTGTGACA	4233
Db	1741	TGTCTCTCACTCAAACTTCTGAAGGAAGATGAAATGACCAAACTGGGACTACGTGTGACA	1800
Qy	4234	CTTGACTCAGATCAGTGTGGCTATCAAGCAGGAGCAATGGCCAGCCCTTCCCTCGCAG	4293
Db	1801	CTTGACTCAGATCAGTGTGGCTATCAAGCAGGAGCAATGGCCAGCCCTTCCCTCGCAG	1860
Qy	4294	TACATCAATGATCTGATAGCGCTTGGTGCGGTGATCCATGGAGGGGCTTCCAGCTT	4353
Db	1861	TACATCAATGATCTGATAGCGCTTGGTGCGGTGATCCATGGAGGGGCTTCCAGCTT	1920
Qy	4354	AGTGAAGGCGGCTTGTCAATGGAATCTATTTATATTTCTGAAAAACATCGTATAA	4410
Db	1921	AGTGAAGGCGGCTTGTCAATGGAATCTATTTATATTTCTGAAAAACATCGTATAA	1977

RESULT 8
AL591398
LOCUS
DEFINITION Human DNA sequence from clone RP11-81C12 on chromosome X, complete sequence.
ACCESSION AL591398
VERSION AL591398.2 GI:14330165
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 89973)
REFERENCE
Heath, P.
Direct Submission
Submitted (05-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

On Jun 8, 2001 this sequence version replaced gi:14133279.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep. This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>. RPI1-81C12 is from the library RPI1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>. VECTER: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RPI1-81C12. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RPI1-81C12 is at 89874 in this sequence. The true right end of clone RPI1-81C12 is at 100 in this sequence. The true right end of clone RPI1-7706 is at 44281 in this sequence.

Location/Qualifiers

source

```

1..89973
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RPI1-81C12"
/clone_lib="RPI1-11.1"
1..791
/note="L1M2 repeat: matches 5524. .6308 of consensus"
914..1079
/note="L1M2 repeat: matches 994. .1160 of consensus"
1130..1329
/note="L1M3 repeat: matches 2218. .2411 of consensus"
1330..1377
/note="L1M3 repeat: matches 318. .365 of consensus"
1378..1680
/note="L1M3 repeat: matches 1. .308 of consensus"
1681..1848
/note="L1M3 repeat: matches 155. .318 of consensus"
1849..2152
/note="L1M3 repeat: matches 1. .299 of consensus"
2153..2305
/note="L1M3 repeat: matches 1. .155 of consensus"
2306..2602
/note="L1M3 repeat: matches 1901. .2218 of consensus"
2642..3475
/note="L1M3 repeat: matches 1245. .2090 of consensus"
3476..3860
/note="L1M3 repeat: matches 3. .371 of consensus"
3861..3929
/note="L1M3 repeat: matches 2090. .2159 of consensus"
3930..4240
/note="L1M3 repeat: matches 1. .306 of consensus"
4241..4885
/note="L1M3 repeat: matches 2159. .2430 of consensus"
4886..5198
/note="L1M3 repeat: matches 1. .309 of consensus"
5199..5491
/note="L1M3 repeat: matches 2430. .2726 of consensus"
5491..5917
/note="L1M3 repeat: matches 3146. .3578 of consensus"
5918..6235
/note="L1M3 repeat: matches 1. .312 of consensus"
6236..6976
/note="L1M3 repeat: matches 3578. .4297 of consensus"
7047..7160
/note="L1M3 repeat: matches 2022. .2140 of consensus"
7161..7262

```

```

repeat_region
/note="L1M3 repeat: matches 7440. .7542 of consensus"
7263..7581
/note="L1M3 repeat: matches 1. .302 of consensus"
7582..7776
/note="L1M3 repeat: matches 7542. .7734 of consensus"
7777..7791
/note="L1M3 repeat: matches 2140. .2153 of consensus"
8810..8960
/note="L1M3 repeat: matches 2610. .2761 of consensus"
9033..9458
/note="L1M3 repeat: matches 2190. .2627 of consensus"
9506..9826
/note="L1M3 repeat: matches 1. .308 of consensus"
9938..10017
/note="L1M3 repeat: matches 1. .80 of consensus"
10019..10166
/note="L1M3 repeat: matches 6611. .6747 of consensus"
10205..10502
/note="L1M3 repeat: matches 1. .311 of consensus"
10538..10681
/note="L1M3 repeat: matches 6468. .6609 of consensus"
11563..11712
/note="L1M3 repeat: matches 13. .165 of consensus"
11906..12226
/note="L1M3 repeat: matches 1. .312 of consensus"
12345..12652
/note="L1M3 repeat: matches 1. .310 of consensus"
12992..13267
/note="L1M3 repeat: matches 1. .285 of consensus"
13268..13327
/note="L1M3 repeat: matches 1. .285 of consensus"
14758..14984
/note="L1M3 repeat: matches 16. .257 of consensus"
15103..15315
/note="L1M3 repeat: matches 1551. .1774 of consensus"
15324..15625
/note="L1M3 repeat: matches 2442. .2750 of consensus"
16017..16079
/note="L1M3 repeat: matches 78. .139 of consensus"
16085..16215
/note="L1M3 repeat: matches 704. .1687 of consensus"
16594..16617
/note="L1M3 repeat: matches 2 mer tg 95% conserved"
16923..16980
/note="L1M3 repeat: matches 2 mer aa 74% conserved"
17294..17457
/note="L1M3 repeat: matches 12. .254 of consensus"
17730..18153
/note="L1M3 repeat: matches 12. .254 of consensus"
17743..17945
/note="L1M3 repeat: matches 12. .254 of consensus"
18255..18278
/note="L1M3 repeat: matches 12. .254 of consensus"
18578..18872
/note="L1M3 repeat: matches 1. .234 of consensus"
19101..19464
/note="L1M3 repeat: matches 13. .388 of consensus"
19508..19598
/note="L1M3 repeat: matches 2662. .2750 of consensus"
20286..20583
/note="L1M3 repeat: matches 1. .293 of consensus"
21186..21309
/note="L1M3 repeat: matches 2 mer tt 59% conserved"
21665..21708
/note="L1M3 repeat: matches 2 mer ac 77% conserved"
23271..23419
/note="L1M3 repeat: matches 37. .173 of consensus"
23420..23720
/note="L1M3 repeat: matches 3. .310 of consensus"
23721..23800
/note="L1M3 repeat: matches 173. .255 of consensus"
23906..23960
/note="L1M3 repeat: matches 394. .448 of consensus"

```


[illegible]

```

QY 4316 CCTGGTCCGCTGATCCATGAGGGGCTCCAGCTTAGTGAGGGCCCGCTGTGTCATGG 4375
DB 53640 CCTGGTCCGCTGATCCATGAGGGGCTGTGAGCTAGTGAGGGCTCTGCTGCTCATTGG 53699
QY 4376 AACTCATCTTTTATATT 4392
DB 53700 AACTCATCTTTTATTTT 53716

RESULT 9
LOCUS AF104305
DEFINITION Xenopus laevis Smad anchor for receptor activation (SARA) mRNA,
complete cds.
ACCESSION AF104305
VERSION AF104305.1 GI:4092768
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
Tsukazaki, T., Chiang, T.A., Davison, A.F., Attisano, L. and Wrana, J.L.
SARA, a FIVE domain protein that recruits Smad2 to the TGFbeta
receptor
Cell 95 (6), 779-791 (1998)
JOURNAL 99081294
MEDLINE 986596
PUBMED
REFERENCE 2 (bases 1 to 4823)
AUTHORS Wrana, J.L.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1998) Program in Developmental Biology, Hospital
for Sick Children, 555 University Avenue, Toronto, Ontario M6G 1X8,
Canada

FEATURES
source
1..4823
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
1..4823
/gene="SARA"
316..4023
/gene="SARA"
/codon_start=1
/product="Smad anchor for receptor activation"
/protein_id="AAC99463.1"
/db_xref="GI:4092769"
/translation="MPKXVIGDRIIMAEDSLFPNNTGSEIVCNSEIVSNESQSLQALDDQS
VNIHNEKSVLLADGFPSCSPKSLINFDCLTMENPLSHQMSVDDNDKETVTSIVLP
TIIQDTSNVDPAINKPGTKPEHRALKETTSVILPEIKPYSTCAALSFPNNKVPYSY
QLNNTDLSVSPVACSEQQKHSTSLHEEKLFEVGSATESFAATAETVLIDNEALR
SAEPFDIVKRNFSQCVINGDLTKCSGLSSENEKFCASKEPFGVDANVLLENACVA
YKBAIDLPEBNGTNAPMSLYNGCDSYGMKNPVAQNPKNLPKSEDSVTBEKETEESSK
EYTVGVVEQREDDVTERGILLNAKADQKNLHLSLCNQVPSMHSQTSPKXKIVQS
LSVPGGARTQPTHLKHPKLPSEMLQSDLIIPNAGGSKYKNDMLKNSQGDNLI
SELRSDSARVSPVTDANGHPPEYRGPSGLCLAVSPDSPNDLLAGQFVPSKPF
TLGEVAPVWPDSQAPNCKCEARFTFKRHECRACGVFCACCSLXKQLQYMDKK
EARCVIACHVSLNAQWENMLASVQSPNPNPABYCSITPPMQQAQSGALSSPPP
TMVPVGLVXHPGTEGQSKEORRVWFADGILPNEGTDSDNANVTVAAGTLTVSHTN
NSTSEBENTSGFGSITQVGSAMNLI PEDGLPILISTGVKGDYAVEERPSQMSVMQ
QLBEGGPDPLAVFLVNLNALLAMKLVNVRKCHCFTTKGMHVGQREIVILQCLPDE
CLPRDLUFHFVLIQAIQANVVGNLHGSFLSQFLGSKDGGFLYVAPTSIQDGL
VLPAEPLVGLILQKMETPWAKVFPFIRLMRLGAERYLPCLPFSVYRKPLFGETH
TINVLADFNYQVTLPEVQGLVVDMEVRSITKIPSNRYNEMKMNKSNHVLAIIG
ACFNQADSHLVQVNDGNYQTQVSIHQPRTKVTGSPFFVPSGALKSSGYLAKSS
IVBEGVMVQITASMDALROSLREMKDFTTCGKADEBSQEHVYVOWVEDDKNFKG
VFSPIDCKSSVTSVKI FHGSEVSKAGKTIIRWIEVFELDNEEQQSGLSDPADHSLRT
ENKAFACIALCHLKLKEDGMTRGLSLVSLDSQVGVQAGSNGQLLPARTNLDLDG
ALVPIVHGTCQQLSEGPVSMELIFYLENIS"

Query Match 31.9%; Score 1545.8; DB 5; Length 4823;

```

```

Best Local Similarity 71.8%; Pred. No. 0;
Matches 2136; Conservative 0; Mismatches 792; Indels 45; Gaps 7;

QY 1443 ACAGTGTGTGTGATTGGCAGATGTCAGGTCTAGATTATAAAGGAACCTGCATTAGTGAAG 1502
DB 1089 AGAGTTTGAAGGAGGTAGTCTAATGTCCTGTGGAATGATGATGTAGTCTATTA 1148
QY 1503 TGAAGAAATGTGATTTCTCCACTGTTATAGACACACAGCAGCAAAATATCTATCTAATGG 1562
DB 1149 AGAAGCAATAGATTTGCTTGAAGAAATGAACTAATGACCAATGCTCTCTGTAATGG 1208
QY 1563 TTGTGATTCCTTGAATGCAAGACCCAGCTGTTCTTTTGTTCGAAGACTTTACCTC 1622
DB 1209 GTGTGATTCCTTGAATGAAACCCAGCCGTAGCTCAAAACCCCAAGAAATTTACCTTC 1268
QY 1623 CAAAGAAGATTGAGTAACAGAGAAAGAAATAGAGGAAAGCAAGTCAGATGTACTC 1682
DB 1269 AAAAGAAGATTCTGTGACAGAGAAAGAAATGAAGAAAGCAAGTCAGATGTACTC 1328
QY 1683 AATATTATG---AACAGAGAGAAATGAGGCCACAGAGGAGTGTGACTTCTTTAAA 1739
DB 1329 TGGTGTATTATGAACAACAAGAGAGATGATGTTACAGAGAGAGTGTGACTTCTGTTAAA 1388
QY 1740 CAGCACTGTGTGACCTTAATGAAGAAATATTATTAATATTTCTGTAGTCAAGTTCATC 1799
DB 1389 TGCTAAGGCTGACCAATGAAGAAACA---ATTGTGATGTCTTTGTAATCAGTTCATC 1445
QY 1800 AGTGTGTGGGCAATCTTCCCAAGGTAGTAGCAAGCTGCAATCTATCAGTGTTCCTTT 1859
DB 1446 CATGATGCGCAACATCACCAAAAGGCAAGATTGTGCAATCTCTCAGTGTTCCTATA 1505
QY 1860 TGTGTGTGCAAGACCCCAAGCACTTCTAATCTTAACTTAACTTCAATTCCTCAAGCCATATC 1919
DB 1506 CGGTGAGCACCACATAGACGCCCAACTCTCAACTCCAACTCCAACTCCAAAGCCATGTC 1565
QY 1920 AGACCAATTTACAAA---ATGACTTTCTGCAAAAGTGTGAATATACTAAAAATAAAAA 1976
DB 1566 TGAATATTGTGAGAGGATCTCATCTCTCAATGCTGGTGCAGCTCTAAATACAAAA 1625
QY 1977 TGATATTCTTGGGAAGCAAAATAGGGGAAACTCAGCAACCAATGTATGCGATCCATC 2036
DB 1626 TGACATGTTTAAACAAATCAAAATCAGGGGGAATAACCTGATTTTCAAGATCACTTGCCTGAGGA 1685
QY 2037 TTTGGGAAACATCTCTAATGTGATACAAATGGGGAACATTTAGAAAGTTATGAGGCTGA 2096
DB 1686 TTCTGAGTGCAGCCGCTGT---TACTGATGCTAATGGTGTATTCCTGCGAGATACAG 1742
QY 2097 GATCTCCACTAGACCATGCTTGCATTAGTCTCCAGATGACCCAGATATGATCTCAGAGC 2156
DB 1743 GGGACCTGGCAGCTTGTGCTTGTGCTGCTCTCCAGACAGCCAGACGATCTGCTTGC 1802
QY 2157 TGGTCAGTTTGGAAATTTCTGCGAGAAAGCCATTACACCGCTGGGTGAGGTGGCTCCAGT 2216
DB 1803 CGGGCAGTTTGGGTACCCATCTCTAAGCCATTTACTACTCTAGGGAGTGTGCTCCAGT 1862
QY 2217 ATGGTACCGGATTTCAAGCTCCAAATGATGAATGTGAACCGAGTTTATCATTCAC 2276
DB 1863 CTGGGTGCGAGATTTCCCAAGCACCACAACTGATGAAGTGCAGGCGCAGATTTACATTTAC 1922
QY 2277 CAAAAGGAGGATCATCTCAGAGCATGTGGGAAGGTTTCTGTGCTTCTCTCTGTAGCT 2336
DB 1923 CAAAAGGAGGATCATCTCAGGAGCTTGTGGGAAGGTTTCTGTGCTTCTGTGCTGCT 1982
QY 2337 GAAATGTAACTGTTTATACATGGAAGAAAGAGTAGAGTGTGTGATCTGCCATTC 2396
DB 1983 AAAATGCAACTACAGATACATGGAATGAAAGAGGCTGCTGTGTGTGTTATTGTCTATC 2042
QY 2397 AGTGTATGATGCTCAAGCTGGGGAACATGATGATGCTCAAGCCAGAGCCCTAA 2456
DB 2043 TGTGTTTATGATGCTCAAGCATGGGAGACATGTTTAAAGTGCATCGGTCCAAAGCCCAA 2102
QY 2457 CCCTAACCAATCTGCTGAATATCTTCTTACTATATCCCTCCCTTGCAGCAAGCTCAGGCTC 2516

```

ORIGIN


```
Qy 1701 AGGAATGAGGCGACAGAGGAGTGACTCTTTTAAACAGCACTGGTGACCTAATGAA 1760
Db 59239 AGGAACAGAAATGCAAGAGGAGTGACTCTTTTAAACAGCACTGGGAATGAA 59180
Qy 1761 GAAAAATATTACATAATTTCTGTAGTCAAGTTCCATCATCATCTGGCAATCTCCCC 1820
Db 59179 GAAAAATATTACATAATTTCTGTAGTCAAGTTCCATCATCATCTGGCAATCTCCCC 59120
Qy 1821 CAAGTAGTAGCAAGCTGGCATCTATCACTGTGTTCTTTTGGTGGTCAAGCCCAAGCA 1880
Db 59119 CAAG--ATAGCAACCTGCAATCTATAGTGTCTTTTGGTGGTCAAGCCCAAGCA 59063
Qy 1881 ACCTTCTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTT 1940
Db 59062 ACCTTCTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTT 59003
Qy 1941 TCCTGCAACAGTGGAAATATATCTAAATAATAATAATAATAATAATAATAATAATAATA 2000
Db 59002 TCCTGCAACAGTGGAAATATATCTAAATAATAATAATAATAATAATAATAATAATAATA 58943
Qy 2001 AGGGGAAACTCAGCAACCAATGATGATGATGATGATGATGATGATGATGATGATGAT 2060
Db 58942 AGGGGAAACTCAGCAACCAATGATGATGATGATGATGATGATGATGATGATGATGAT 58883
Qy 2061 TACAATGGGGAACATTTGAAAGCTTATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 2120
Db 58882 TACAATGGGGAACATTTGAAAGCTTATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 58823
Qy 2121 ATTAGCTCCAGATAGCCAGCAATATGATCTCAGAGCTGGTCAAGTTTGGAAATTTG 2180
Db 58822 ACTAGCTCCAGATAGCCAGCAATATGATCTCAGAGCTGGTCAAGTTTGGAAATTTG 58763
Qy 2181 AAAGCCATTCACAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2240
Db 58762 AAAGCCATTCACAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 58703
Qy 2241 AAATGTGATGAATGTGAAGCAGCTTACATTCACCAAAAGGAGGATCACTCAGAGC 2300
Db 58702 AAATGTGATGAATGTGAAGCAGCTTACATTCACCAAAAGGAGGATCACTCAGAGC 58643
Qy 2301 ATGTGGGAAGGT 2312
Db 58642 GTGTGGGAAGGT 58631

RESULT 11
AC095988/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-29P9, *** SEQUENCING IN PROGRESS ***
AC095988
AC095988.7 GI:30522870
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
Rattus
1 (bases 1 to 258787)
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alibrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A., Chu, J.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, B.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished
2 (bases 1 to 258787)
Worley, K.C.

Direct Submission

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 258787)
Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:24818127.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GEC
Center clone name: CH230-29P9
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 248170 bases at least Q40

```

Consensus quality: 250504 bases at least Q30
Consensus quality: 252206 bases at least Q20
Estimated insert size: 262793; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 254666: contig of 254666 bp in length
*
* 254667 254766: gap of unknown length
*
* 254767 255900: contig of 1134 bp in length
*
* 255901 256000: gap of unknown length
*
* 256001 258787: contig of 2787 bp in length.
*
FEATURES             location/Qualifiers
source               1..258787
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-29P9"
misc_feature         1..1729
                    /note="wgs_end_extension"
                    clone_end:T7"
                    7543..8441
                    /note="clone boundary"
                    clone_end:T7"
                    site:EcoRI
                    end_sequence:BH261874"
misc_feature         1011..102472
                    /note="clone boundary"
                    clone_end:Sp6
                    site:EcoRI
                    end_sequence:BH261875"
misc_feature         250578..254666
                    /note="wgs_end_extension"
                    clone_end:Sp6"

ORIGIN
Query Match      28.9%; Score 1399.2; DB 2; Length 258787;
Best Local Similarity 81.3%; Pred. No. 0;
Matches 1718; Conservative 0; Mismatches 313; Indels 81; Gaps 5;
QY 201 AGATGAACACAGTTCTTCTTCTTATTTGGATACAAAGTGAATTAAGATTTAGATCCCC 260
DB 182539 AGATGAACACAGTTCTTCTTCTTATTTGGATACAAAGTGAATTAAGATTTAGATTTCTC 182480
QY 261 TTCTCACCGGTGTGTCATTTAACTTACCTTCTTGGCCAGTGTGAATGAATCTGCAGTTTCTAA 320
DB 182479 TTCTATCATTTGTTTAACTTACCTTCTTGGATACAAAGTGAATTAAGATTTAGATTTCTAA 182420
QY 321 TGATGTCAACCAACCACTGAAGTCTTCTCCTGGCTCATTCAGCTCCCTGACCAACAGA 380
DB 182419 TGACACAGGACACAGCGGAAATCTTCTTCTTGGCTGCTAGCTCCCTTACAGGGA 182360
QY 381 GGAAGAGATCAGTGTCTATGACAGGACTGTAACTTAATCCAGAGATTCACCAAT 440
DB 182359 GGGAAAGGTCCTGTGCTACAGCAGGACGGTAGTCTTAATCCAGAGATTCACCAAT 182300
QY 441 GTGATTTGATGAAATGCTGTGTCAGAGACCACTTAATTAAGAGAACTATATAGTTGGGA 500
DB 182299 GTGATTTGATGAAATGCTGTGTCAGAGACCACTTAATTAAGAGAACTATACAACTCAGGA 182240
QY 501 TGATCAATGAGTCTGTTGAGTGGGAGAGAGAAATGTGGAAACCTGGCTTGTCTGCC 560
DB 182239 TGATCAATTTAGTCCGTTGAAGTGGGAGAGAGAAACGTGGGAGGCTCACTTGTCTGCC 182180

```

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.G.M., Gabris, A., Gancu, R., Garcia, A., Garner, T., Garza, M.A., Gebregregorius, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hughes, M., Hollans, B., Howells, S., Huiyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshuhea, L., Loutsegged, H., Lozador, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwaekemele, O., Okwuonu, G., Olarnpunsgoon, A., Pal, S., Parks, K., Pasternack, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Platterner, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, J., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shwartzbeyn, A., Sisson, I., Sitter, C.D., Smales, D., Sneed, A., Sodargren, E., Song, X.-Z., Soreller, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

Direct Submission
Unpublished

2 (bases 1 to 201330)
Worley, K.C.

Direct Submission
Submitted (13-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 201330)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Nov 13, 2002 this sequence version replaced gi:21319493.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly ('contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

Center project name: GVBD
Center clone name: CH230-324B23


```

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 188149 bases at least Q40
Consensus quality: 190201 bases at least Q30
Consensus quality: 191458 bases at least Q20
Estimated insert size: 194406; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  * consists of 2 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
*
* 1 198699: contig of 198699 bp in length
* 198700 198799: gap of unknown length
* 198800 201330: contig of 2531 bp in length.
FEATURES
      source
        1..201330
          /organism="Rattus norvegicus"
          /mol_type="genomic DNA"
          /db_xref="taxon:10116"
          /clone="CH230-324B23"
      misc_feature
        1..1448
          /note="wgs end_extension"
          clone_end:Sp6"
          1920..4036
      misc_feature
        /note="wgs end_extension"
          clone_end:Sp6"
      misc_feature
        complement(6631..7552)
          /note="clone boundary"
          clone_end:Sp6
          site:
          end_sequence:B2238666"

ORIGIN
Query Match      26.2%; Score 1265.6; DB 2; Length 201330;
Best Local Similarity 77.3%; Pred. No. 1.2e-301;
Matches 1632; Conservative 0; Mismatches 398; Indels 82; Gaps 6;

QY 201 AGATGAACAGTTCTTCTACTTTATGATGACAAAGTGAATAGATTCTAGATCCCC 260
DB 35153 AGATGAACAGTTCTTCTACTTTATGATGACAAAGTGAATAGATTCTAGATCCCTC 35094
QY 261 TTCTCACCGGTGTCATTTAACCCCTACTTTGGCCAGTGTGAATGAATCTGCAGTTCTAA 320
DB 35093 TTCTCATCATTGTCTTTTAAACCCCTGCATTAGCCAAATGTGAATGAACCTACAATTTCTAA 35034
QY 321 TGAGTCACAAACACAACTGAAGTCTTCCCTGGCTCATTCAGTCCCTGCACACAGA 380
DB 35033 TGAGCAGGACACAGCGGAAATCTTTTCTTGTGGTCTAGTCCCTTTACAAGGA 34974
QY 381 GGAAGAGGATCAGTGTGTAATGACAGAGTGAATCTAAATCCAGAGATTGCCACAAT 440
DB 34973 GGGAAAGGTCCCTGTGCTAACGGACAGACGGTAGTCTAAATCCAGAGATTGACACGAT 34914
QY 441 GTGGATTGATGAAATGCTGTGTGAGAGACAGTGAATTAAGAGAACTATAGTTGGGA 500
DB 34913 GTGGATTGATGAAATGCTGTGGCAGAGACCACTAATTAAGAGAACTACAAATCAGGA 34854
QY 501 TGATCAATGCAAGTGTGTTGAAGTGGGAGAGAAATGTGGAACCTGGCTGTCTGCC 560
DB 34853 TGATCAATTTAGTCCGTTGAAGTGGGAGAGAACTGAGAGCTCTACTGTCTGCC 34794
QY 561 AGATGAGAAAGTGTCTTGTGTAGCCGTCATGCAATCTGATGATTAAGAGCAATPACA 620
DB 34793 AGACGAGAAGATGTTCTTGTGTAGTGTGTCATGCAATCTGATGATTAAGAGCACTGCA 34734
QY 621 AAACGATTACAGGATTGTAATAATTAATAGTCAATCCCTTATGGATCCTTTTAGCTG 680

```

```

Db 34733 AAGTGATTTTCAGGATTGTAAATAACTATAATAGTCAATCCCTCATGATTTCTTTAGCTG 34674
QY 681 TTCACCTGGATAAATAAGACAGACAAACTGATCAATTTAGTTAGTATAAATAGTCCAC 740
Db 34673 TTCACCTGGATAAATAAGACAGACAAACTGATCAATTTAGTTAGTATAAATAGTCCAC 34614
QY 741 TGAAGAAGATATCAATTTACAGAGAAACAAATGGAATCCATTTGAATAGACCCGAAACAGAGG 800
Db 34613 AGAAAGGTATTAATCTCAGAGAAACAAATGGATGCTTTGAATTAACCAATACAGAG 34554
QY 801 GAGATCTGTTAAACCATCTGTCTCTACTTCACTGATCTGATCTAGCCAGTGTCTGTCC 860
Db 34553 GGAATCTCTTAATCATCTCTGCGCTCTCTTCAAGTAAACAGTCAACCATCTCTCTC 34494
QY 861 TTCACAAATTAAGAGTACGGAAGTATAGTACAGACCCCTCCATCTCTGCGATTACAAG 920
Db 34493 TTCACAAATTAAGAGTACGGAAGTATAGTACAGACCCCTCTACATCTACAGTACAAG 34434
QY 921 TTTAACGGTTGATTCAGTAAATCTCATCCAGGAAACAGATGATGCTCTGTTTAAAAA 980
Db 34433 CTTAGCAATTAAT-----ACATCCAGGGAATGGATGGAGGCTTGCATTTAAAAA 34383
QY 981 GCAAGAGACTATATACAGATGAGGACCTCACTGGCAAAATCAGCTCTCTAGGACAGA 1040
Db 34382 ACAAGAGATTATATGCTGTATGAGACCTTTCTGGCGCAAAACAGCTCTCTAGGACAGA 34323
QY 1041 TCTAGGAGTCCAAATTTCTTTTCCACATGATGAGGGGATTTTGAATGAAAAAGAGCC 1100
Db 34322 TCTAGGAGTCCAAATTTCTTTTCCACATGAGGGGAGGCTTTAAACAAAAAGAGCC 34263
QY 1101 ASCAGAGGAGACCACTGAGAAATCCCTCGGTCTGTTTACCTTCTCTCAAC 1160
Db 34262 AGCAGAGGAGACAGCTGAAACCCCTCCATCTGTTTATCTTTGAATCTCAAGCC 34203
QY 1161 AGACATGCCTTAATGGGTCTGGAAGGAATTAATGACTGTGAACGGTGTTCAGATTGCTGT 1220
Db 34202 AGATACTCTCTGTTTATGTGGAAGGACAAT---TCCGAACCATCTCAGACTGCTGT 34146
QY 1221 GCCTAATGAAGTTAGGGCTGATGAAATGAAGTTTATGAACATGAAGAACTCTTGGCAC 1280
Db 34145 GTCTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 34103
QY 1281 TACAGAAATCTTAAATATACAGAGCAATTTCTCTGAATCTCAGGACATGACTAATTTGAA 1340
Db 34102 -----CAGACGACTACTGGA 34086
QY 1341 GTTGACTAAACTAAATGAGATGAATGATGATGATGATGATGATGATGATGATGATGAT 1400
Db 34085 CTGAGCAAAAGTAAATGAGTGGATGATGATGATGATGATGATGATGATGATGATGAT 34026
QY 1401 ACAGATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1460
Db 34025 ACAGATCAGCAACCTTGAGGACCTCA-----GAGTGCAGGAGAGTGTGTAGATGCG 33972
QY 1461 AGATGAGGTCTAGATTTAAAAAGGAATTTGATGATGATGATGATGATGATGATGATGAT 1520
Db 33971 AGATTCAGATCTAGATTTCAAAGGAATTTGATGATGATGATGATGATGATGATGATGAT 33912
QY 1521 CACTGTATAGACACACAGGACCAATTTATCTATCTAATGTTGTTGTTGTTGTTGTTGTTG 1580
Db 33911 CACTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 33852
QY 1581 GCAAGACCCAGGTGTTCTTTTGTTCGAAAGCTTTTACCTCCAAAGAGATTCAGTAAAC 1640
Db 33851 GCAGAGCCCAATTTGATCTTTTGTTCGAAAGCTTTTACCTCCAAAGAGATTCAGTAAAC 33792
QY 1641 AGAAGAAAAAGAAATAGAGGAAAGCAAGTCAAGATGCTCAATATTTATGAACAGAG 1700
Db 33791 AGAAGAAAAAGAAATAGAGGAAAGCAAGTCAAGATGCTCAATATTTATGAACAGAG 33732
QY 1701 AGGAATGAGGACACAGAGGAGTGGACTACTTTTAAACAGCAGCTGGTGAATCAATCAA 1760

```


TITLE	Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
JOURNAL	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16999-16903 (2002)
REFERENCE	22388257
PUBMED	12477932
2 (bases 1 to 1237)	
STRAUSBERG, R.	
DIRECT SUBMISSION	
TITLE	Submitted (31-JAN-2003) National Institutes of Health, Mammalian
JOURNAL	Gene Collection (MGC), Cancer Genomics Office, National Cancer
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
	USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgapbs-remail.nih.gov Tissue Procurement: Jeffrey E. Green, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc_mgc@nhgri.nih.gov Akter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, O.L., Masiello, C., Mastri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantropop, S., Thomas, P.J., Touchman, J.W., Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Series: IRAP Plate: 100 Row: 1 Column: 8

This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction.

FEATURES	Location/Qualifiers
Source	1..1237 /organism="Mus musculus" /mol_type="mRNA" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:6477813" /tissue_type="Colon, normal. 5 month old male mouse." /clone_fib="NCI CGAP_Co24" /lab_host="DH105" /note="Vector: PCMV-SPORT6"
gene	<1..1237 /gene="Madhip" /note="synonyms: SARA, ZFYV99" /db_xref="LocusID:230597" /db_xref="MGI:2652838"
CDS	<1..806 /gene="Madhip" /codon_start=3 /product="Madhip protein" /protein_id="AAH46341.1" /db_xref="GI:28374416" /db_xref="LocusID:230597" /db_xref="MGI:2652838" /translation="LAGACAFCEKADSHLVQVNDGNYQTOAISTHNPCKVTGASF FVFSALKSSGSLAKSSIVGQWQVITAEKNDLSRLQALRMKDTITCGKADSSG NDHHIOWDDDKTHQVNVKGVSPIDGKSSITNVKIFGSEYKANGKVIWTEVFLE NDDHNCILSDPADHSLTEHVAKFCALCPHLKLIKDKGMTKGLRVLDSQVQVIG AGSGQPLSPQYMWNLDSALVPVHGACQLSEGPVVMVELIFILENA"

ORIGIN	Query Match 18.8%; Score 898.8; DB 10; Length 1237; Best Local Similarity 87.2%; Pred. No. 6.7e-211;
--------	---

Matches 1080; Conservative 0; Mismatches 142; Indels 16; Gaps 8;

QY 3605 TCCTGCGAGAGTCTCTCTCAATGAAGAGCAGACTCTCATCTCTGCTGTGTACAGA 3664
 Db 1 TACTGCGAGAGTCTCTCTCAATGAAGAGCAGACTCTCATCTCTGCTGTGTACAGA 60

QY 3665 ATGATGATGGAATATCAGACCCAGGCTATCAGTATTCACAATCAGCCGAGAAAGTGA 3724
 Db 61 ACATGATGGAATATCAACCCAGGCTATCAGTATTCACAATCAGCCGAGAAAGTGA 120

QY 3725 CTGGTCCAGTCTCTGTTTCAGTGGGCTCTGAATCTCTTCTGATACCTTGCCA 3784
 Db 121 CTGGGCGCAGTCTCTGTTTCAGTGGGCTCTGAATCTCTTCTGATACCTTGCCA 180

QY 3785 AGTCCAGTATTCGGAAGATGTTGTTATGGTCCAGATTACTGCAGAGAACATGGATTCCT 3844
 Db 181 AGTCCAGTATTCAGAGATGAGTATGATGCTGCTCCAGATCACTGGGAGAACATGGATTCCT 240

QY 3845 TGAGGAGGCTCTGCGAGAGATGAAGGACTTCACCATCCTGTGGGAGGCGGAGCGG 3904
 Db 241 TGAGGAGGCTCTGCGGAGATGAAGGACTTTACCATCCTGTGGGAGGCGGAGCGG 300

QY 3905 AGGAACCCAGGAGCACATCCATCCAGTGGTGTGATGACAGAAACGTTAGCAGG 3964
 Db 301 AGGACCCAGGAGCACATCCATCCAGTGGTGTGATGACAGAAACGTTAGCAGG 360

QY 3965 GTGTCGTAAGTCTATAGATGGGAAGTCCATGGAGCTATTAACAAATGTGAAGATATCC 4024
 Db 361 GTGTCGTAAGTCTATAGATGGGAAGTCTATGGATCTTATAACAAACGTGAAGATATCC 420

QY 4025 ATGGATCAGATATTAAGCAATGAAGAAATGAAATATCAGATGACAGAGTGTTCCTAG 4084
 Db 421 ATGGGTGAGATATTAAGCAATGAAGAAATGAAATATCAGATGACAGAGTGTTCCTAG 480

QY 4085 AAAACGATCAGCAGACCAATTCCTCAGTGTCTGAGATCAGATGATGATGATGATG 4144
 Db 481 AAAACGATCAGCAGACCAATTCCTCAGTGTCTGAGATCAGATGATGATGATGATG 540

QY 4145 ATGTTGCCAAGCTTTTGGCTTGTCTCTGCTCTCAGTCTGATGATGATGATGATG 4204
 Db 541 ATGTTGCCAAGCTTTTGGCTTGTCTCTGCTCTCAGTCTGATGATGATGATGATG 600

QY 4205 GAATGACCAATCTGGAGTACGTGTGACATCTGATCAGATCAGTGTGCTATCAAGCAG 4264
 Db 601 GAATGACCAATCTGGAGTACGTGTGACATCTGATCAGATCAGTGTGCTATCAAGCAG 660

QY 4265 GGAGCAATGGCAGCCCTTCCCTGCGAGTACATGAATGATCTGGATAGCCCTTGCTGC 4324
 Db 661 GGAGCAATGGCAGCCCTTCCCTGCGAGTACATGAATGATCTGGATAGCCCTTGCTGC 720

QY 4325 CGGTGATCCATGGAGGCGCTGCGAGCTTGTGAGGCGCCGTTGTGATGAGTCTATCT 4384
 Db 721 CGGTGATCCATGGAGGCGCTGCGAGCTTGTGAGGCGCCGTTGTGATGAGTCTATCT 780

QY 4385 TTTATATCTGAAACATCTGATATAACAGAGAGACTTCAATTTTTCTGTCAGACTT 4444
 Db 781 TTTATATCTGAAACATCTGATATAACAGAGAGACTTCAATTTTTCTGTCAGACTT 836

QY 4445 GTTGCAGCAGCAGTATACCAATATCTTTCATCTTAAACTGGAATTAAGCTTAACT 4503
 Db 837 GTTGCAGCAGCAGTATACCAATATCTTTCATCTTAAACTGGAATTAAGCTTAACT 896

QY 4504 GTTAACACTATTATGGGCTGGGAAATAGGTTGGGAGTGGGGTTTGGGAGACGGTGGG 4563
 Db 897 GTTAACACTATTATGGGCTGGGAAATAGGTTGGGAGTGGGGTTTGGGAGACGGTGGG 956

QY 4564 AAAGGTGTTGGGGGACCAATGTTTCAATATCTTAAGTCTTCTATGCAATGTCACCA 4623
 Db 957 AAAGGTGTTGGGGGACCAATGTTTCAATATCTTAAGTCTTCTATGCAATGTCACCA 1011

QY 4624 AGAAGATCTGGGAGCTTCTGCTCTGCAACAGTATGATGATCTCTGAGGCTATCC 4683
 Db 1012 AGAAGATCTGGGAGCTTCTGCTCTGCAACAGTATGATGATCTCTGAGGCTATCC 1069

QY 4684 CTCTGTTACTGTTTAGACAAGATTCGCTCTCTCTCAAGATTTACTTATGGTCACTG 4743
 Db 1070 CTCTGTTACTGTTTAGACAAGATTCGCTCTCTCTCTCTATGA-TTACTTCCGTCATGT 1128

QY 4744 GTCAGAA--AATGCTCAAAATGGGTACAAACCATCACCAGGGTGGGATGGAGGCGAGAG 4801
 Db 1129 GTCAGATGAATGCTCATATGGGCACCAACCATCACCAGGTCAATGGG-GGGCAGAAA 1187

QY 4802 GGAATATAATATAAGCATCAAAAARAAAAA 4839
 Db 1188 GGAATATAATATAAGCATCGTCAAAAARAAAAA 1225

RESULT 14
 AC068614 164706 bp DNA linear PRI 09-JAN-2002
 LOCUS Homo sapiens BAC clone RP11-512N1 from 2, complete sequence.
 DEFINITION AC068614
 ACCESSION AC068614.5 GI:15145605
 VERSION HTG.
 KEYWORDS SOURCE Homo sapiens (human)

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 (bases 1 to 164706)
 Sultston,J.E. and Waterston,R.
 Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074

REFERENCE
 2 (bases 1 to 164706)
 Kang,K., Abbott,A., Boyer,E. and Dixon,R.
 The sequence of Homo sapiens BAC clone RP11-512N1
 Unpublished (2001)

REFERENCE
 3 (bases 1 to 164706)
 Waterston,R.H.
 Direct Submission

TITLE
 Submitted (05-MAY-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE
 4 (bases 1 to 164706)
 Waterston,R.H.
 Direct Submission

TITLE
 Submitted (09-AUG-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE
 5 (bases 1 to 164706)
 Waterston,R.H.
 Direct Submission

TITLE
 Submitted (09-JAN-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Aug 9, 2001 this sequence version replaced gi:14628397.

COMMENT
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: sapiens@wustl.edu
 ----- Summary Statistics
 ----- Center project name: H_NH0512N01

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by

restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Oseegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Taten, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-102G8, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-512N1; actual end is at base position 164706 of RP11-512N1.

Polymorphisms exist between RP11-512N1 and RP11-102G8. Data from AC032014 was used to finish this clone AC068614.

FEATURES

source

1..164706

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone="RP11-512N1"

/clone_lib="RPCI-11"

1..138

/rpt_family="MER1_type"

145..804

/rpt_family="L1"

836..1569

/rpt_family="L1"

1685..1848

/rpt_family="MER1_type"

1868..2044

/rpt_family="MER1_type"

2260..2411

/rpt_family="CR1"

2485..3182

/rpt_family="L2"

3172..3293

/rpt_family="L2"

3298..3488

/rpt_family="L2"

3458..3484

/rpt_family="AT-rich"

3617..3718

/rpt_family="MER1_type"

3719..3886

/rpt_family="Alu"

3887..3946

/rpt_family="MER1_type"

3954..4018

/rpt_family="MIR"

5044..5194

/rpt_family="MIR"

5700..5780

/rpt_family="MER1_type"

6317..6551

/rpt_family="MIR"

6824..6871

/rpt_family="CT-rich"

7157..7256

/rpt_family="L2"

8291..8715

/rpt_family="L2"

/rpt_family="L2"

/rpt_family="L2"

/rpt_family="L2"

/rpt_family="L2"

/rpt_family="L2"

/rpt_family="L2"

/rpt_family="L2"

/rpt_family="L2"

/rpt_family="L2"

/rpt_family="L2"

/rpt_family="L2"

/rpt_family="L2"

/rpt_family="L2"

repeat_region
8744..8909
/rpt_family="MaLR"
repeat_region
8933..8971
/rpt_family="L2"
repeat_region
9722..9786
/rpt_family="MIR"
repeat_region
10431..10489
/rpt_family="L2"
repeat_region
10686..10897
/rpt_family="L2"
repeat_region
12174..12338
/rpt_family="MIR"
repeat_region
12430..12505
/rpt_family="L1"
repeat_region
12898..12963
/rpt_family="L2"
repeat_region
13850..13945
/rpt_family="Alu"
repeat_region
14692..15044
/rpt_family="MIR"
repeat_region
15938..16906
/rpt_family="L2"
repeat_region
17595..18092
/rpt_family="L1"
repeat_region
18081..18507
/rpt_family="AchoBo"
repeat_region
18374..18426
/rpt_family="AchoBo"
repeat_region
19734..19816
/rpt_family="AT-rich"
repeat_region
19971..20124
/rpt_family="MIR"
repeat_region
20810..20877
/rpt_family="MER103"
repeat_region
21098..21164
/rpt_family="(TGA)n"
repeat_region
21865..22233
/rpt_family="L2"
repeat_region
22328..22439
/rpt_family="MIR"
repeat_region
22440..22664
/rpt_family="MIR"
repeat_region
22665..22747
/rpt_family="MIR"
repeat_region
23939..24091
/rpt_family="MIR"
repeat_region
24668..24701
/rpt_family="(TTTA)n"
repeat_region
24783..25084
/rpt_family="Alu"
repeat_region
25063..25084
/rpt_family="AT-rich"
repeat_region
25255..25297
/rpt_family="L2"
repeat_region
26227..27186
/rpt_family="ERV1"
repeat_region
27187..27284
/rpt_family="ERV1"
repeat_region
27285..27610
/rpt_family="ERV1"
repeat_region
27848..28178
/rpt_family="L2"
repeat_region
29215..29249
/rpt_family="AT-rich"
repeat_region
30350..30386
/rpt_family="AT-rich"
repeat_region
32688..32732
/rpt_family="MIR"
repeat_region
33373..33665
/rpt_family="MIR"

Query Match 18.1%; Score 875.4; DB 9; Length 164706;
Best Local Similarity 90.2%; Pred. No. 3.8e-205;

source 1. .1144
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:5011916"
/tissue type="Mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(clonal) outgrowth infected with the virus MMV.""
/clone_lib="NCI_GCAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

Query Match 15.3%; Score 740.8; DB 10; Length 1144;
Best Local Similarity 85.7%; Pred. No. 8.1e-172; Indels 16; Gaps 8;
Matches 919; Conservative 0; Mismatches 137;

QY 3755 CTCTGAATCCTCTCTTGATACCTTGCACAGTCCAGTATTGTGGAAGATGGTGTATGG 3814
Db 10 CGCTGAAATCCTCTCTAGATACCTTGCACAGTCCAGTATTGTGGAAGATGGATCATGG 69

QY 3815 TCCAGATTACTGCAGAGACATGATTCCTTTGAGGAGGACCTGCGAGAGATGAAGACT 3874
Db 70 TCCAGATCACAGCGGAGAACATGACTCTTTGAGGAGGACCTGCGGAGATGAAGACT 129

QY 3875 TCACCATCACCTGTGGGAGGCGAGCGAGGAAACCCAGGAGCACATCCACATCCAGT 3934
Db 130 TTACCATCACCTGTGGGAGGCGGATGCAGAGGACCCCGAGAACAGATCCACATCCAGT 189

QY 3935 GGGTGGATGATGACAGAACGTTTACAGAGGTGCTGTAAGTCTTATAGATGGGAAGTCCA 3994
Db 190 GGGTGGATGATGACAGAACGTTTACAGAGGTGCTGTAAGTCTTATAGATGGGAAGTCTA 249

QY 3995 TGGAGACTATACAAATGTGAAGTATTTCCTAGATGATGAGATATAAAGCAATGCAAAAG 4054
Db 250 TGGAACTTATACAAACGTTGAATATTTCCTAGATGATGAGATATAAAGCAATGCAAAAG 309

QY 4055 TAATCAGATGGACAGAGGTGTTTTCTTAGAAAAAGATGACACAGACACAAATTCGCTCAGTG 4114
Db 310 TCATCAGATGGACAGAGGTATTTTCTTGGAAAAAGATGACACACAAATTCGCTCAGTG 369

QY 4115 ATCTGCGAGATCACAGTATGATGAGTATGCTGATGATGCTGCAAGGCTTTTGCCTTCTCTCT 4174
Db 370 ACCCTGCGAGATCACAGTATGATGAGTATGCTGATGATGCTGCAAGGCTTTTGCCTTCTCT 429

QY 4175 GTCTTCACTGAACTTCTGAAGGAGATGGAATGACCAAACTGGGACTAGCTGTGACAC 4234
Db 430 GCCCCCACTGAGCTTTTAAAGAGATGGAATGACAAACTGGGACTCGTGTGACGC 489

QY 4235 TTGACTCAGATCAGTGGCTTATCAGCAGGAGCAATGGCCAGCCCTTCCCTCGCAGT 4294
Db 490 TTGACTCAGATCAGTGGCTTATCAGCAGGAGCAATGGCCAGCCCTTCCCTCGCAGT 549

QY 4295 ACATGAATGATCTGGATAGCGCTTGTGCGGCTGATCCATGGAGGGGCTGCCAGGCTTA 4354
Db 550 ACATGAACGATCTGACAGTGGCTTGTGCGCTGCTGCTTATGAGGAGGCTGCCAGGCTCA 609

QY 4355 GTGAGGGCCCGTGTCTGATGGAATCATCTTTTATATTTCGAAAAACATCGTATAAACAG 4414
Db 610 GTGAGGGCCCTGTCTGATGGAATCATCTTTTATATTTCGAAAAACATCGCAT 665

QY 4415 AGAAGACTTCATTTTTCTGTTTCAGACTGTTGCAACAGCAGTATACCCCAATCATTT 4474
Db 666 ATGAGGACTCCATTTTCTGTTTCAGACTGTTGCAACAGCAGTATACCCCAATCATTT 725

QY 4475 GCACTTTAAACTGGAAGATTAAAGC-TTTTGTAACTATTAAATGGGGTGGGAATAGG 4533
Db 726 GCACTTTAAACTGGAAGATTAAAGCTTTTGTAACTATGTCAGTAGGAGGGTTTGGGG 785

QY 4534 GTGGAGTGGGGGTTTGGGAGCGGTGGGAAGGGTGGTTGGGGGGGACCGATGTTCCAT 4593
Db 786 TGGGAATAGGACTTAGGGAGATGGGTGGGAGGGTGGT--AGGGGACAGATGTTCCAT 843

Search completed: May 18, 2004, 02:10:13
Job time : 12033.2 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 09:46:50 ; Search time 1114.1 seconds

(without alignments)

18451.753 Million cell updates/sec

Title: us-09-744-167-1

Perfect score: 4839

Sequence: 1 gcatactgaatcagcaggac.....atcaaaaaaaaaaaaaaaaa 4839

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373663 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4839	100.0	4839	3	Aaz50068 Human Sma
2	1977	40.9	1977	9	Adc64443 Human SAR
3	1830	37.8	1830	5	Aas30091 Human lun
4	1830	37.8	1830	5	Aas30092 Human lun
5	1830	37.8	1830	9	Adb33428 Human nov
6	1830	37.8	1830	9	Adb33429 Human nov
7	1545.8	31.9	4823	3	Aaz50070 Xenopus S
8	699.4	14.5	888	7	Aca56987 Human adi
9	576	11.9	576	7	ACA57333 Human adi
10	570.4	11.8	610	5	Aas29903 Human lun
11	570.4	11.8	610	9	Adb33148 Human nov
12	560.6	11.6	6632	3	Aaz50069 Human Sma
13	559	11.6	5402	6	Abk11709 DNA encod
14	333.2	6.9	2678	3	Aaz50071 Xenopus S
15	318	6.6	332	8	Ach30390 Human sec
16	260.2	5.4	498	8	ACH22774 Human adu
17	245.8	5.1	4068	4	Abi28217 Drosophil
18	220.8	4.6	969	6	Abi18932 Human pol
19	203.4	4.2	475	8	ACH16512 Human adu
20	202	4.2	6784	4	Abi28216 Drosophil
21	160	3.3	164	3	ACa07133 Human sec
22	154	3.2	437	7	ABX54300 Bovine ES
23	143.4	3.0	316	9	ADB56467 Toxicity

24	118	2.4	118	5	AAS30093	Human lun
25	118	2.4	118	9	ADB33430	Human nov
26	89.2	1.8	391	8	ACH20104	Human adu
27	63.8	1.3	701	6	ABQ57538	Human col
28	60.8	1.3	289	2	AAV88493	EST clone
29	60	1.2	60	6	ABN41947	Human spl
30	54.8	1.1	1639	5	AAS67754	DNA encod
31	53	1.1	396	3	AAH31021	Human col
32	53	1.1	472	3	AACT7594	Human ORF
33	53	1.1	479	6	ABO99035	Human ORF
34	53	1.1	489	6	ABSG9469	Novel mur
35	53	1.1	731	9	ADC32237	Human nov
36	53	1.1	977	9	ADC30462	Human nov
37	53	1.1	1362	6	AAH17982	Human CDN
38	53	1.1	1434	5	AAH23820	Human tra
39	52.4	1.1	1296	6	ABK49898	Human CDN
40	52.4	1.1	1821	4	AAH15014	Human CDN
41	52.4	1.1	3610	6	ABNS9933	Novel hum
42	51.6	1.1	2000	7	ADA71398	Rice gene
43	50	1.0	476	8	ACH30322	Human tes
44	50	1.0	2138	2	AAZ22280	Human tes
45	50	1.0	2138	8	ACD40524	CDNA sequ

ALIGNMENTS

RESULT 1

AAZ50068

ID AAZ50068 standard; cDNA; 4839 BP.

XX AC AAZ50068;

XX DT 04-MAY-2000 (first entry)

XX DE Human Smad Anchor for Receptor Activation protein-1 encoding cDNA.

XX KW Smad Anchor for Receptor Activation protein; hSARA1; human;

KW transforming growth factor-beta; TGF-beta; bone morphogenetic protein;

KW BMP; activin; anti-inflammatory; cytotatic; antiarthritic; vulnary;

KW TGF-beta modulator; wound healing; scarring; arthritis; immune response;

KW inflammatory response; tumour progression; cell proliferation; fibrosis;

KW fibrogenesis; tissue morphogenesis; tissue damage; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 439..4410

XX FT /*tag= a

XX FT /product= "Human SARA-1 protein"

XX FT /note= "Binds to receptor regulated Smad proteins"

XX FT /transl_except= (pos:4177..4179, aa:Thr)

XX FT /transl_except= (pos:4180..4182, aa:Gln)

XX FT /transl_except= (pos:4198..4200, aa:Gly)

XX FT /transl_except= (pos:4279..4281, aa:His)

XX FT /transl_except= (pos:4306..4308, aa:Phe)

XX FT /transl_except= (pos:4315..4317, aa:Asp)

XX FT /transl_except= (pos:4324..4326, aa:Lys)

XX FT /transl_except= (pos:4327..4329, aa:Met)

XX WO200005360-A1.

XX 03-FEB-2000.

XX 20-JUL-1999; 99WO-CA000656.

XX 20-JUL-1998; 98CA-02237701.

XX 10-DEC-1998; 98CA-02253647.

XX (HSCR-) HSC RES & DEV LP.

XX Wrana JL;

XX

QY 1741 AGCACTGGTGAACCTAATGAAGAAAAATTAATTAATAAATTCGTAGTCAAGTTCATCA 1800
DB 1741 AGCACTGGTGAACCTAATGAAGAAAAATTAATTAATAAATTCGTAGTCAAGTTCATCA 1800
QY 1801 GTGCTGGGCAATCTCCCAAGAGTAGTAGAGAGCGTCCATCTCATAGTGTCTCTTT 1860
DB 1801 GTGCTGGGCAATCTCCCAAGAGTAGTAGAGAGCGTCCATCTCATAGTGTCTCTTT 1860
QY 1861 GGTGGTCAAGACCAAGCAACCTCTAATCTTAAACTTCAAAATCCAAAGCCATTATCA 1920
DB 1861 GGTGGTCAAGACCAAGCAACCTCTAATCTTAAACTTCAAAATCCAAAGCCATTATCA 1920
QY 1921 GACCAATTAATAAATGACTTCTCGTCAAAAGTGGAAATTAATTAATAAATAAATATGAT 1980
DB 1921 GACCAATTAATAAATGACTTCTCGTCAAAAGTGGAAATTAATTAATAAATAAATATGAT 1980
QY 1981 ATTCTTGGGAAGCAAAATTAGGGGAAACTCAGCAACCAATGTATGAGTCCATCTTG 2040
DB 1981 ATTCTTGGGAAGCAAAATTAGGGGAAACTCAGCAACCAATGTATGAGTCCATCTTG 2040
QY 2041 GGAACAATCTTAATGTCGATCAAAATGGGAAACATTTAGAAAGTTATGAGGCTGAGATC 2100
DB 2041 GGAACAATCTTAATGTCGATCAAAATGGGAAACATTTAGAAAGTTATGAGGCTGAGATC 2100
QY 2101 TCCACTAGACCATGCTTGCATTAAGTCCAGATAGCCAGATTAATGATCTCAGAGCTGGT 2160
DB 2101 TCCACTAGACCATGCTTGCATTAAGTCCAGATAGCCAGATTAATGATCTCAGAGCTGGT 2160
QY 2161 CAGTTTGGAAATTTCTGCCAGAAAGCCATTACCAACGCTGGGTGAGTGGTCCAGTATGG 2220
DB 2161 CAGTTTGGAAATTTCTGCCAGAAAGCCATTACCAACGCTGGGTGAGTGGTCCAGTATGG 2220
QY 2221 GTACCGGATTTCTCAGGCTCAAAATGTGATGAATGTGAAGCCAGGTTTACATTCACCAA 2280
DB 2221 GTACCGGATTTCTCAGGCTCAAAATGTGATGAATGTGAAGCCAGGTTTACATTCACCAA 2280
QY 2281 AGGAGGATCACTCAGAGATGTGGAAAGTCTTCTGCTCTCCCTGCTAGCCTGAAA 2340
DB 2281 AGGAGGATCACTCAGAGATGTGGAAAGTCTTCTGCTCTCCCTGCTAGCCTGAAA 2340
QY 2341 TGTAAACTGTGTATACATGGACAGAAAGAGCTAGAGTGTGTATCTGCCATTCAGTG 2400
DB 2341 TGTAAACTGTGTATACATGGACAGAAAGAGCTAGAGTGTGTATCTGCCATTCAGTG 2400
QY 2401 CTAAATGATGCTCAAGCTGGGAAACATGATGAGTGCCTCAAGCCAGAGCCCTTAACCT 2460
DB 2401 CTAAATGATGCTCAAGCTGGGAAACATGATGAGTGCCTCAAGCCAGAGCCCTTAACCT 2460
QY 2461 AACAACTCTGCTGAATACTGTCTACTATCCCTCTTGCAAGAGCTCAGGCTCAGGA 2520
DB 2461 AACAACTCTGCTGAATACTGTCTACTATCCCTCTTGCAAGAGCTCAGGCTCAGGA 2520
QY 2521 GCTCTGAGCTTCCACCTCCACCTGTGATGTGATGAGTGGAGTTTAAAGCCCTTGA 2580
DB 2521 GCTCTGAGCTTCCACCTCCACCTGTGATGTGATGAGTGGAGTTTAAAGCCCTTGA 2580
QY 2581 GCAGAAGTGGCTCAGCCAGAGACAGAGCGAGTTTGGTTTGTGATGGATCTTGCC 2640
DB 2581 GCAGAAGTGGCTCAGCCAGAGACAGAGCGAGTTTGGTTTGTGATGGATCTTGCC 2640
QY 2641 AATGGAGAAGTTGCTGATGAGCCAAATTAACAATGAATGGAATCTCTGAGGAAAC 2700
DB 2641 AATGGAGAAGTTGCTGATGAGCCAAATTAACAATGAATGGAATCTCTGAGGAAAC 2700
QY 2701 CTGCTGTGTACAGGACCCAGTCAAGCCAGTACTACCTCTACAGCAGAGAGCG 2760
DB 2701 CTGCTGTGTACAGGACCCAGTCAAGCCAGTACTACCTCTACAGCAGAGAGCG 2760
QY 2761 GATATTTGCTATTTCTCTGGGAGTAACTCAGGTGTGAAGTCTCTGTTGGAAGTGCATG 2820
DB 2761 GATATTTGCTATTTCTCTGGGAGTAACTCAGGTGTGAAGTCTCTGTTGGAAGTGCATG 2820

QY 2821 AATCTTATCTCGAAGATGGCTTCTCCCAATTCATCTCATCTCCACTGGTGTAAAGGAC 2880
DB 2821 AATCTTATCTCGAAGATGGCTTCTCCCAATTCATCTCATCTCCACTGGTGTAAAGGAC 2880
QY 2881 TATGCTGTGGAAGAGAAACCATCACAGATTTTCAAGTAATGACAGAGTTGAGAGATGGTGGC 2940
DB 2881 TATGCTGTGGAAGAGAAACCATCACAGATTTTCAAGTAATGACAGAGTTGAGAGATGGTGGC 2940
QY 2941 CCTGACCCACTGTATTTTGTATTAATGCAAAATTTGTTGTCAATGGTTTAAATTTGTAAT 3000
DB 2941 CCTGACCCACTGTATTTTGTATTAATGCAAAATTTGTTGTCAATGGTTTAAATTTGTAAT 3000
QY 3001 TATGTGAACAGAGTGTGTTTCAACCAAGGCAATGTCATGAGTGGTGGTCACTCT 3060
DB 3001 TATGTGAACAGAGTGTGTTTCAACCAAGGCAATGTCATGAGTGGTGGTCACTCT 3060
QY 3061 GAGATAGTCAATCTTCTACAGTGTTTACCGGATGAAAGTGTGTTGCCAAAGATATCTTT 3120
DB 3061 GAGATAGTCAATCTTCTACAGTGTTTACCGGATGAAAGTGTGTTGCCAAAGATATCTTT 3120
QY 3121 AATCACTTTGTCAGCTTTATCGGATGCTCTGGCAGGGAATGTTGTCAGCAACTGGGA 3180
DB 3121 AATCACTTTGTCAGCTTTATCGGATGCTCTGGCAGGGAATGTTGTCAGCAACTGGGA 3180
QY 3181 CATTCCTTTCTCAGTCAAAAGTTCCTTGGCAGTAAAGAACATGTTGATATGATG 3240
DB 3181 CATTCCTTTCTCAGTCAAAAGTTCCTTGGCAGTAAAGAACATGTTGATATGATG 3240
QY 3241 ACATCTACCTACAGTCAAGACCTAGTACTCCCAACCCACCTTACTTGTGGG 3300
DB 3241 ACATCTACCTACAGTCAAGACCTAGTACTCCCAACCCACCTTACTTGTGGG 3300
QY 3301 ATTCTTATCCAGAAATGGAAACTCTTGGGCTAAAGTATTTCCCTATCCGCTGATGTTG 3360
DB 3301 ATTCTTATCCAGAAATGGAAACTCTTGGGCTAAAGTATTTCCCTATCCGCTGATGTTG 3360
QY 3361 AGACTTGGAGCTGAATATCGACTTTATCCATGCCACTATTCACTGTCAGATTTGCGAAG 3420
DB 3361 AGACTTGGAGCTGAATATCGACTTTATCCATGCCACTATTCACTGTCAGATTTGCGAAG 3420
QY 3421 CCATTTGTTGGAGAGCGGGCATACCATCATGATTTCTTGGCAGACTTCAGAAATATAC 3480
DB 3421 CCATTTGTTGGAGAGCGGGCATACCATCATGATTTCTTGGCAGACTTCAGAAATATAC 3480
QY 3481 CAGTATACCTTCCCAAGTGTTCAGAGTTTGGTGTGATATGGAAGTTTGGAAAACTAGC 3540
DB 3481 CAGTATACCTTCCCAAGTGTTCAGAGTTTGGTGTGATATGGAAGTTTGGAAAACTAGC 3540
QY 3541 ATCAAAATTTCCAGCAACAGATACATGAGATGATGAAGCCATGAACAAGTCCCAATGAG 3600
DB 3541 ATCAAAATTTCCAGCAACAGATACATGAGATGATGAAGCCATGAACAAGTCCCAATGAG 3600
QY 3601 CATGCTCGCAGGAGTGCCTGCTTCAATGAAAGGCGAGACTCTCATCTTGTGTGTA 3660
DB 3601 CATGCTCGCAGGAGTGCCTGCTTCAATGAAAGGCGAGACTCTCATCTTGTGTGTA 3660
QY 3661 CAGAAATGATGTAAGAACTATCAGACCCAGGCTATCAGTATTCACAATCAGCCCAAGAAA 3720
DB 3661 CAGAAATGATGTAAGAACTATCAGACCCAGGCTATCAGTATTCACAATCAGCCCAAGAAA 3720
QY 3721 GTGACTGTGCGAGTTTCTTGTGTTTCTGAGGCTCTGAAATCTCTTCTGATACCTT 3780
DB 3721 GTGACTGTGCGAGTTTCTTGTGTTTCTGAGGCTCTGAAATCTCTTCTGATACCTT 3780
QY 3781 GCCAAGTCCAGTATTTGTGAAGATGGTGTATGTTGTCAGATTTACTCGAGAGAACATGGAT 3840
DB 3781 GCCAAGTCCAGTATTTGTGAAGATGGTGTATGTTGTCAGATTTACTCGAGAGAACATGGAT 3840
QY 3841 TCCTTGAGCAGGCACTGCGAGAGATGAAGGATTCACATCAGCTGTGGAGCGGAC 3900
DB 3841 TCCTTGAGCAGGCACTGCGAGAGATGAAGGATTCACATCAGCTGTGGAGCGGAC 3900
QY 3901 GCGGAGGAACCCAGGAGCACATCCACATCCAGTGGGTGATGATGACAAAGACGTTAGC 3960

QY	2854	CTCATCTCCACTGGTGTAAAGGAGACTATGCTGTGGAAGAGAAACCATCAGAGTTTCA	29113
Db	421	CTCATCTCCACTGGTGTAAAGGAGACTATGCTGTGGAAGAGAAACCATCAGAGTTTCA	480
QY	2914	GTAAAGCAGAGTTGAGAGAGTGGCCCTGACCCACTTTGTATTTGTTTAAATGCAAT	2973
Db	481	GTAAAGCAGAGTTGAGAGAGTGGCCCTGACCCACTTTGTATTTGTTTAAATGCAAT	540
QY	2974	TTGTTGTCATGTTAAATTTGTAATATATGTGAACAGGAAGTCTGTTGTTTCACAAC	3033
Db	541	TTGTTGTCATGTTAAATTTGTAATATATGTGAACAGGAAGTCTGTTGTTTCACAAC	600
QY	3034	AAGGGAATGCATCCAGTGGGTGAGTCTGAGATAGTCAATCTTCTACAGTGTTTACCGAT	3093
Db	601	AAGGGAATGCATCCAGTGGGTGAGTCTGAGATAGTCAATCTTCTACAGTGTTTACCGAT	660
QY	3094	GAAGAAGTTTGCAAGGATATCTTTAATCACTTTGTCAGCTTTTATCGGATGCTCTG	3153
Db	661	GAAGAAGTTTGCAAGGATATCTTTAATCACTTTGTCAGCTTTTATCGGATGCTCTG	720
QY	3154	GCAGGGAATGTTGGTGGACATCTTCCCTTCTTCCAGTCAAAAGTTTCTTTGSCAGT	3213
Db	721	GCAGGGAATGTTGGTGGACATCTTCCCTTCTTCCAGTCAAAAGTTTCTTTGSCAGT	780
QY	3214	AAAGACATGGTGGATCTTTATATGTGACATCTACCTACCAGTCACTGCAAGACCTTAGTA	3273
Db	781	AAAGACATGGTGGATCTTTATATGTGACATCTACCTACCAGTCACTGCAAGACCTTAGTA	840
QY	3274	CTCCCAACCCACTTACTTTGTTGGGATCTTTATCCAGAAATGGGAAATCTCTTTGGGCT	3333
Db	841	CTCCCAACCCACTTACTTTGTTGGGATCTTTATCCAGAAATGGGAAATCTCTTTGGGCT	900
QY	3334	AAAGTATTTCCATTCCTGCTGATGTTTGAGACTTTGGAGCTGAATATGCACTTTATCCATGC	3393
Db	901	AAAGTATTTCCATTCCTGCTGATGTTTGAGACTTTGGAGCTGAATATGCACTTTATCCATGC	960
QY	3394	CCACTATTCAGTGTCAATTTCCGAAGCCATTTGTTGGAGAGAGCGGGCATACCATCATG	3453
Db	961	CCACTATTCAGTGTCAATTTCCGAAGCCATTTGTTGGAGAGAGCGGGCATACCATCATG	1020
QY	3454	AATCTTCTTGCGACTTTCAGAAATACCAAGTATACCTTGCAGTAGTTCAGGTTTGGTG	3513
Db	1021	AATCTTCTTGCGACTTTCAGAAATACCAAGTATACCTTGCAGTAGTTCAGGTTTGGTG	1080
QY	3514	GTTGATATGAAGTTCCGAAAATAGCATCAAAATTCGCCAGCAACAGATACATGAGATG	3573
Db	1081	GTTGATATGAAGTTCCGAAAATAGCATCAAAATTCGCCAGCAACAGATACATGAGATG	1140
QY	3574	ATGAAGCCATGAACAAGTCCAATGAGCATGCTCTGCGAGAGGTGCTGCTTCAATGAA	3633
Db	1141	ATGAAGCCATGAACAAGTCCAATGAGCATGCTCTGCGAGAGGTGCTGCTTCAATGAA	1200
QY	3634	AAGCAGACTCTCATCTTTGTGTGTGATGATGATGGAACCTATCAGACCCAGGCT	3693
Db	1201	AAGCAGACTCTCATCTTTGTGTGTGATGATGATGGAACCTATCAGACCCAGGCT	1260
QY	3694	ATCAGTATTCACATCAGCCAGAAAAGTGACTGGTCCAGTTTCTTTGTTGTTCAAGTGC	3753
Db	1261	ATCAGTATTCACATCAGCCAGAAAAGTGACTGGTCCAGTTTCTTTGTTGTTCAAGTGC	1320
QY	3754	GCTCTGAAATCCTCTTCTGGATACCTTTGCCAAGTCCAGTATGTGGAAGATGTTGTTTATG	3813
Db	1321	GCTCTGAAATCCTCTTCTGGATACCTTTGCCAAGTCCAGTATGTGGAAGATGTTGTTTATG	1380
QY	3814	GTCAGATTACTGCAGAGAACATGGATTCCTTTGAGGAGGACCTGCAGAGATGAAGGAC	3873
Db	1381	GTCAGATTACTGCAGAGAACATGGATTCCTTTGAGGAGGACCTGCAGAGATGAAGGAC	1440
QY	3874	TTCCACCATCACTGTGGGAAGGGCGAATCCGAGGAAACCCAGGAGCACATCCCATCCAG	3933
Db	1441	TTCCACCATCACTGTGGGAAGGGCGAATCCGAGGAAACCCAGGAGCACATCCCATCCAG	1500

Qy	3934	TGGGTGATGATGACAAAGACGTTAGCAAGGTTGCTGAAGTCCTTAGATGGAGTCC	3999
Db	1501	TGGGTGATGATGACAAAGACGTTAGCAAGGTTGCTGAAGTCCTTAGATGGAGTCC	1560
Qy	3994	ATGAGAGCTATAACAAATGTGAAGATATTCCATGGGATCAGAAATATAAGCAAAATGGAAA	4053
Db	1561	ATGGAGACTATACAAATGTGAAGATATTCCATGGGATCAGAAATATAAGCAAAATGGAAA	1620
Qy	4054	GTAATCAGATGGACAGAGTGTGTTTTCTTAGAAAAAGATGACAGACCAATTCGCTCAGT	4113
Db	1621	GTAATCAGATGGACAGAGTGTGTTTTCTTAGAAAAAGATGACAGACCAATTCGCTCAGT	1680
Qy	4114	GATCCTGCAGATCACAGTAGATTGACTGAGCATGTTGCCAAAGCTTTTTCGCTTCTCTC	4173
Db	1681	GATCCTGCAGATCACAGTAGATTGACTGAGCATGTTGCCAAAGCTTTTTCGCTTCTCTC	1740
Qy	4174	TGTCCTCACCTGAAACTTCTCAAGGAAGATGGAAATGACCAAACTGGGACTACGTGTGACA	4233
Db	1741	TGTCCTCACCTGAAACTTCTCAAGGAAGATGGAAATGACCAAACTGGGACTACGTGTGACA	1800
Qy	4234	CTTGACTCAGATCAGGTTGGCTATCAAGCAGGAGGAGCAATGCCAGCCCTTCCCTCGCAG	4293
Db	1801	CTTGACTCAGATCAGGTTGGCTATCAAGCAGGAGGAGCAATGCCAGCCCTTCCCTCGCAG	1860
Qy	4294	TACATGAATGATCTGGATAGCGCTTGGTGC CGGTGATCCATGAGAGGGGCTGCCAGCTT	4353
Db	1861	TACATGAATGATCTGGATAGCGCTTGGTGC CGGTGATCCATGAGAGGGGCTGCCAGCTT	1920
Qy	4354	AGTGAGGGCCCCGTTGTCATGGAACCTCATCTTTTATATCTCGAATAACATCGTATAA	4410
Db	1921	AGTGAGGGCCCCGTTGTCATGGAACCTCATCTTTTATATCTCGAATAACATCGTATAA	1977
RESULT 3			
AAS30091			
ID	AAS30091 standard; DNA; 1830 BP.		
XX			
AC	AAS30091;		
XX			
DT	21-NOV-2001 (first entry)		
XX			
DE	Human lung antigen genomic DNA #161.		
XX			
KW	Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;		
KW	chicken; sheep; immunosuppressive; antiarthritic; vasotropic;		
KW	antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;		
KW	cerebroprotective; nontropic; antibacterial; virucide; fungicide; cancer;		
KW	ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;		
KW	hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;		
KW	cerebrovascular disorder; nervous system disorder; bacterial infection;		
KW	fungal infection; viral infection; ocular disorder; endocrine disorder;		
KW	gastrointestinal disorder; renal disorder; respiratory disorder;		
KW	wound healing; skin aging; organ transplantation; food preservative;		
KW	tissue regeneration; anti-infertility; food additive.		
OS	Homo sapiens.		
XX			
PN	WO200155303-A2.		
XX			
PD	02-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-US0001301.		
XX			
PR	31-JAN-2000; 2000US-0179065P.		
PR	04-FEB-2000; 2000US-0180628P.		
PR	24-FEB-2000; 2000US-0184664P.		
PR	02-MAR-2000; 2000US-0186350P.		
PR	16-MAR-2000; 2000US-0189874P.		
PR	17-MAR-2000; 2000US-0190078P.		
PR	18-APR-2000; 2000US-0198123P.		
PR	19-MAY-2000; 2000US-0205515P.		
PR	07-JUN-2000; 2000US-0209467P.		
PR	28-JUN-2000; 2000US-0214886P.		

disorders include autoimmune diseases such as rheumatoid arthritis, hyperproliferative disorders such as neoplasms of the breast or liver, cardiovascular disorders such as cardiac arrest, cerebrovascular disorders such as cerebral ischaemia, nervous system disorders such as Alzheimer's disease, infections caused by bacteria, viruses and fungi, ocular disorders such as corneal infection, endocrine disorders such as premature labour and infertility, gastrointestinal disorders such as Crohn's disease, renal disorders such as glomerulonephritis and respiratory disorders such as asthma and pleurisy. The polypeptides can also be used to aid wound healing, to prevent skin aging due to sunburn, to maintain organs before transplantation, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Query Match 37.8%; Score 1830; DB 5; Length 1830;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1830; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

201 AGATGAACAGTCTCTCTACTTTATTTGGATACAAAGTGGATPAAGATTCTAGATCCCC 260
1 AGATGAACAGTCTCTCTACTTTATTTGGATACAAAGTGGATPAAGATTCTAGATCCCC 60

261 TTCTCAGCGCTGTCAATTAACCCCTACTTTGGCCAGTGTGAATGAATCTGCAGTTCTAA 320
61 TTCTCAGCGCTGTCAATTAACCCCTACTTTGGCCAGTGTGAATGAATCTGCAGTTCTAA 120

321 TGAGTCAACACCACTGAAAGTCTTCTCCCTGGCTCATTACAGTCCCTCCACACAGA 380
121 TGAGTCAACACCACTGAAAGTCTTCTCCCTGGCTCATTACAGTCCCTCCACACAGA 180

381 GGAAGAGGATCACTGTCTAATGAGCAGAGTCTGTAAATCTAAATCCAGAGATTGCCAAT 440
181 GGAAGAGGATCACTGTCTAATGAGCAGAGTCTGTAAATCTAAATCCAGAGATTGCCAAT 240

441 GTGAGTGTGAAGTCTCTTCCAGAGACAGTCTAATTAAGAGAACTATAGTTGGGA 500
241 GTGAGTGTGAAGTCTCTTCCAGAGACAGTCTAATTAAGAGAACTATAGTTGGGA 300

501 TGATCAATGAGTGTCTTGAAGTGGAGAGAAATGTGAAACCTGGCTGTCTGCC 560
301 TGATCAATGAGTGTCTTGAAGTGGAGAGAAATGTGAAACCTGGCTGTCTGCC 360

561 AGATGAGAAGTGTCTTGTGTAGCCGTATGATCACTGTGATAAAGAGCATTACA 620
361 AGATGAGAAGTGTCTTGTGTAGCCGTATGATCACTGTGATAAAGAGCATTACA 420

621 AAACGATTTACAGGATGTAAATTAATATAGTCAATCCCTTATGGATGCTTTTAGCTG 680
421 AAACGATTTACAGGATGTAAATTAATATAGTCAATCCCTTATGGATGCTTTTAGCTG 480

681 TTCACTGGATAAAGAAACAGACAACTGATCAATTTAGTTTATAGTATTAATGAGTCCAC 740
481 TTCACTGGATAAAGAAACAGACAACTGATCAATTTAGTTTATAGTATTAATGAGTCCAC 540

741 TGAABAGATATGAATTCAGAGAACAAATGGATCCATTCGAATAGACCGAAACAGAGG 800
541 TGAABAGATATGAATTCAGAGAACAAATGGATCCATTCGAATAGACCGAAACAGAGG 600

801 GAGATCTGTAAACCATCTGTCTCTACTTTCATCTGATAGTCTAGCCAGTGTCTGTCCCC 860
601 GAGATCTGTAAACCATCTGTCTCTACTTTCATCTGATAGTCTAGCCAGTGTCTGTCCCC 660

861 TTCAATTAAGAGTACCGAGTATAGTAGAGCCCTCCATGCTCTCGGATTAACAG 920
661 TTCAATTAAGAGTACCGAGTATAGTAGAGCCCTCCATGCTCTCGGATTAACAG 720

921 TTTAACGGTGTGATTCAGTAACTCTCATCCAGGAAACAGATGGATGTCCTGCTGTAAAAA 980
721 TTTAACGGTGTGATTCAGTAACTCTCATCCAGGAAACAGATGGATGTCCTGCTGTAAAAA 780

991 GCAAGAGAACTATATACAGATGAGGACCTCAGTGGCAAAATCAGCTCTCTTAGACAGA 1040
781 GCAAGAGAACTATATACAGATGAGGACCTCAGTGGCAAAATCAGCTCTCTTAGACAGA 840
1041 TCTAGGAGTCCAAATTCCTTTTCCCATGAGTGGGATTTTGTATGAAAAAGAGCC 1100
841 TCTAGGAGTCCAAATTCCTTTTCCCATGAGTGGGATTTTGTATGAAAAAGAGCC 900
1101 AGCAGAGAGAGCACCCTCGGCTCTGGTTTACCTTTGCTCTCAAAAC 1160
901 AGCAGAGAGAGCACCCTCGGCTCTGGTTTACCTTTGCTCTCAAAAC 960
1161 AGCAGAGAGAGCACCCTCGGCTCTGGTTTACCTTTGCTCTCAAAAC 1220
961 AGCAGAGAGAGCACCCTCGGCTCTGGTTTACCTTTGCTCTCAAAAC 1020
1221 GCCTAATGAAGTTAGGCTGATGAAATGAAGTTTATGAAATGAAGAACTCTTGGCAC 1280
1021 GCCTAATGAAGTTAGGCTGATGAAATGAAGTTTATGAAATGAAGAACTCTTGGCAC 1080
1281 TACAGATTCCTTAATATGACAGAGCATTCTCTGAAATCTCAGGACATGATTAATGGAA 1340
1081 TACAGATTCCTTAATATGACAGAGCATTCTCTGAAATCTCAGGACATGATTAATGGAA 1140
1341 GTTGACTAACTAAATGAGATGAATGATAGCAAGTAAACGAAAGAAAGAAAGTTTCT 1400
1141 GTTGACTAACTAAATGAGATGAATGATAGCAAGTAAACGAAAGAAAGAAAGTTTCT 1200
1401 ACAGATTAAGTCAAGCTGAGGACATTAATGTTGATAGTGGAGACAGTGTGTGGATTGGC 1460
1201 ACAGATTAAGTCAAGCTGAGGACATTAATGTTGATAGTGGAGACAGTGTGTGGATTGGC 1260
1461 AGATCAGAGTCTAGATTTAAAGGAACTTGCATTAGTGAAGTGAAGTGTGATTTCTC 1520
1261 AGATCAGAGTCTAGATTTAAAGGAACTTGCATTAGTGAAGTGAAGTGTGATTTCTC 1320
1521 CACTGTTATAGACACACAGCAGCAAAATATCTATCTAATGTTGTGATTTCTATGGAAT 1580
1321 CACTGTTATAGACACACAGCAGCAAAATATCTATCTAATGTTGTGATTTCTATGGAAT 1380
1581 GCAAGACCCAGGTGTTCTTTTGTTCGAAAGCTTTACCTCCGAAAGAGATTCAGTAAC 1640
1381 GCAAGACCCAGGTGTTCTTTTGTTCGAAAGCTTTACCTCCGAAAGAGATTCAGTAAC 1440
1641 AGAAGAAAAAGAAATAGAGGAAAGCAAGTCAAGATGCTACTCAAAATATTTATGAACAG 1700
1441 AGAAGAAAAAGAAATAGAGGAAAGCAAGTCAAGATGCTACTCAAAATATTTATGAACAG 1500
1701 AGAAGAAAAAGAAATAGAGGAAAGCAAGTCAAGATGCTACTCAAAATATTTATGAACAG 1760
1501 AGAAGAAAAAGAAATAGAGGAAAGCAAGTCAAGATGCTACTCAAAATATTTATGAACAG 1560
1761 GAAAAATTTATTTACATAATTTCTGTAGTCAAGTTCATGCTGGCAATCTTCCCC 1820
1561 GAAAAATTTATTTACATAATTTCTGTAGTCAAGTTCATGCTGGCAATCTTCCCC 1620
1821 CAAAGTAGTAGCAAGCCTGCCATCTATCAGTGTTCCTTTTGGTGTGCAAGCCCAAGCA 1880
1621 CAAAGTAGTAGCAAGCCTGCCATCTATCAGTGTTCCTTTTGGTGTGCAAGCCCAAGCA 1680
1881 AGCTTCTAATCTTAACTTCAAAATTCGAAAGCCATTATCAGACCATTTTACAAATGACTT 1940
1681 AGCTTCTAATCTTAACTTCAAAATTCGAAAGCCATTATCAGACCATTTTACAAATGACTT 1740
1941 TCCTGCAAAACAGTGGAAATTAATCTAAAAATGATATTTCTTGGGAAAGCAAAAT 2000
1741 TCCTGCAAAACAGTGGAAATTAATCTAAAAATGATATTTCTTGGGAAAGCAAAAT 1800
2001 AGGGGAAACTCAGCAACCAATGTATGAG 2030
1801 AGGGGAAACTCAGCAACCAATGTATGAG 1830

RESULT 4
AAS30092
ID AAS30092 standard; DNA; 1830 BP.
XX
AC AAS30092;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human lung antigen genomic DNA #162.
XX
KW Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasia;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; food preservative;
KW tissue regeneration; anti-infertility; food additive.
XX
OS Homo sapiens.
XX
XX
XX WO200155303-A2.
XX
XX
PD 02-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US001301.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0198874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216860P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239353P.
PR 13-OCT-2000; 2000US-0239372P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.

PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249255P.
PR 17-NOV-2000; 2000US-0249256P.
PR 17-NOV-2000; 2000US-0249257P.
PR 17-NOV-2000; 2000US-0249258P.
PR 17-NOV-2000; 2000US-0249259P.
PR 17-NOV-2000; 2000US-0249260P.
PR 17-NOV-2000; 2000US-0249261P.
PR 17-NOV-2000; 2000US-0249262P.
PR 17-NOV-2000; 2000US-0249263P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249266P.
PR 17-NOV-2000; 2000US-0249267P.
PR 17-NOV-2000; 2000US-0249268P.
PR 17-NOV-2000; 2000US-0249269P.
PR 17-NOV-2000; 2000US-0249270P.
PR 17-NOV-2000; 2000US-0249271P.
PR 17-NOV-2000; 2000US-0249272P.
PR 17-NOV-2000; 2000US-0249273P.
PR 17-NOV-2000; 2000US-0249274P.
PR 17-NOV-2000; 2000US-0249275P.
PR 17-NOV-2000; 2000US-0249276P.
PR 17-NOV-2000; 2000US-0249277P.
PR 17-NOV-2000; 2000US-0249278P.
PR 17-NOV-2000; 2000US-0249279P.
PR 17-NOV-2000; 2000US-0249280P.
PR 17-NOV-2000; 2000US-0249281P.
PR 17-NOV-2000; 2000US-0249282P.
PR 17-NOV-2000; 2000US-0249283P.
PR 17-NOV-2000; 2000US-0249284P.
PR 17-NOV-2000; 2000US-0249285P.
PR 17-NOV-2000; 2000US-0249286P.
PR 17-NOV-2000; 2000US-0249287P.
PR 17-NOV-2000; 2000US-0249288P.
PR 17-NOV-2000; 2000US-0249289P.
PR 17-NOV-2000; 2000US-0249290P.
PR 17-NOV-2000; 2000US-0249291P.
PR 17-NOV-2000; 2000US-0249292P.
PR 17-NOV-2000; 2000US-0249293P.
PR 17-NOV-2000; 2000US-0249294P.
PR 17-NOV-2000; 2000US-0249295P.
PR 17-NOV-2000; 2000US-0249296P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249298P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;
WPI; 2001-457723/49.

Isolated polypeptide for treating, preventing and/or prognosing respiratory disorders related to the lung including lung cancers and also for testing and detection e.g. diagnosis.

Claim 1; SEQ ID NO 356; 507pp; English.

The sequences AAS29931-AAS30164 represent genomic DNA molecules, which encode the lung antigen polypeptides of the invention. Lung antigen polypeptides and their associated polynucleotides are useful in the diagnosis, treatment and prevention of various types of disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A pathological condition can be determined by detecting the presence or absence of a mutation in a lung antigen polynucleotide. The treatable disorders include autoimmune diseases such as rheumatoid arthritis, hyperproliferative disorders such as neoplasms of the breast or liver, cardiovascular disorders such as cardiac arrest, cerebrovascular disorders such as cerebral ischaemia, nervous system disorders such as Alzheimer's disease, infections caused by bacteria, viruses and fungi, ocular disorders such as corneal infection, endocrine disorders such as premature labour and infertility, gastrointestinal disorders such as Crohn's disease, renal disorders such as glomerulonephritis and respiratory disorders such as asthma and pleurisy. The polypeptides can also be used to aid wound healing, to prevent skin aging due to sunburn, to maintain organs before transplantation, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Query March 37.8%; Score 1830; DB 5; Length 1830;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1830; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY 201 AGATGAACAGGTTCTTCTACTTTATTGGATAACAAGTGGAATAAGATTCTAGATCCCCC 260
DB 1 AGATGAACAGGTTCTTCTACTTTATTGGATAACAAGTGGAATAAGATTCTAGATCCCCC 60

OY 261 TTCTCACCGGTCTCATTTAAACCCTACTTTGCCAGTGTGAATGAATCTGCAGTTTC TAA 320
DB 61 TTCTCACCGGTGT CATTTAACCCCTACTTTGCCAGTGTGAATGAATCTGCAGTTTC TAA 120

OY 321 TGAGTCACAAACCAACAACTGAAGTCTTCTCCCTGGCTCATTTAGCTCCCCTGCCACAGA 380
DB 121 TGAGTCACAACTCATTAAGTAGAGTTCTCTTGACTCATTTAGCTCCCCTGCCACAGA 180

Db 1261 AGATCAGGCTCTAGATTTTAAAGGAAGTTCGATTAGTGAAGATGATGATTTCTC 1320
QY 1521 CACTGTTATAGACACACAGCAGCAAAATATCTACTAATGTTGTTGATTCCTATGGAAT 1580
Db 1321 CACTGTTATAGACACACAGCAGCAAAATATCTACTAATGTTGTTGATTCCTATGGAAT 1380
QY 1581 GAAGACCCAGGTGTTCTTTGTTTCCAAAGACTTTACCTCCAAAGAAAGATTGATTAAC 1640
Db 1381 GAAGACCCAGGTGTTCTTTGTTTCCAAAGACTTTACCTCCAAAGAAAGATTGATTAAC 1440
QY 1641 ACAAGAAAAGAAATAGAGGAAGCAAGTCAAGTCTACTCAATATTTATGAACAGAG 1700
Db 1441 AGAAGAAAAGAAATAGAGGAAGCAAGTCAAGTCTACTCAATATTTATGAACAGAG 1500
QY 1701 AGGAATGAGGCCACAGAGGAGTGGACTACTTTTAAACAGCACTGTTGGCAATCTTCCC 1760
Db 1501 AGGAATGAGGCCACAGAGGAGTGGACTACTTTTAAACAGCACTGTTGGCAATCTTCCC 1560
QY 1761 GAAAAATATTACATAATTTCTGTAGTCAAGTTCATTCAGTCTTGGCAATCTTCCC 1820
Db 1561 GAAAAATATTACATAATTTCTGTAGTCAAGTTCATTCAGTCTTGGCAATCTTCCC 1620
QY 1821 CAAGTAGTAGCAAGCTGCCATCTATCAGTGTTCCTTTTGGTGGTCAAGACCAAGCA 1880
Db 1621 CAAGTAGTAGCAAGCTGCCATCTATCAGTGTTCCTTTTGGTGGTCAAGACCAAGCA 1680
QY 1881 ACCTTCTAATCTTAACTTCAAAATCCAAAGCCATTTATCAGACCAATTTACAAATGACTT 1940
Db 1681 ACCTTCTAATCTTAACTTCAAAATCCAAAGCCATTTATCAGACCAATTTACAAATGACTT 1740
QY 1941 TCCTGCAACAGTGAATATATCTATAAATAAATAATCTTCTGGGAAGCAAAAT 2000
Db 1741 TCCTGCAACAGTGAATATATCTATAAATAAATAATCTTCTGGGAAGCAAAAT 1800
QY 2001 AGGGAAAACCTCAGCAACCAATGATGCAG 2030
Db 1801 AGGGAAAACCTCAGCAACCAATGATGCAG 1830
RESULT 5
ADB33428
ID ADB33428 standard; DNA; 1830 BP.
XX
AC ADB33428;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human novel lung related polypeptide DNA SEQ ID NO 355.
XX
KW gene therapy; lung antigen; neoplasia; acute myelogenous leukaemia;
KW adenocarcinoma; respiratory disorder; chronic rhinitis; sinusitis;
KW immunodeficiency; X-linked agammaglobulinemia;
KW X-linked infantile agammaglobulinemia; inflammatory disorder;
KW adrenitis; alveolitis; immune complex disease; serum sickness;
KW polyarthritis nodosa; bleeding disorder; thrombocytopenia;
KW Von Willebrand's disease; acquired platelet dysfunction; kidney failure;
KW multiple myeloma; macrophage related disorder; Gaucher's disease;
KW Neimann-Pick disease; tumour; colon cancer; pancreatic cancer;
KW renal disorder; nephritis; bone disorder; Albers-Schonberg disease;
KW bowleg; muscle disorder; Becker's muscular dystrophy;
KW Duchenne's muscular dystrophy; nervous disorder; ischaemic lesion;
KW traumatic lesion; endocrine disorder; Cushing's syndrome;
KW corticosteroid deficiency; gastrointestinal disorder; dysphagia;
KW gastric reflux; human; ds.
OS Homo sapiens.
XX
FN US2003054368-A1.
XX
PD 20-MAR-2003.
XX
XX 22-FEB-2002; 2002US-00079854.
XX
XX

XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.

PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0241826P.
 PR 08-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249298P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 06-DEC-2000; 2000US-0256719P.
 PR 08-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 PR 17-JAN-2001; 2001US-00754878.
 PR (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM, Barash SC;
 XX WPI; 2003-695900/66.
 XX Novel isolated lung antigen polypeptides useful for treating, preventing,
 PT

PT diagnosing acute myelogenous leukemias, adenocarcinoma, thrombocytopenia,
 PT Von Willebrand's disease.
 XX Disclosure; SEQ ID NO 355; 178pp; English.
 XX The invention relates to an isolated lung antigen polypeptide sequence or
 CC encoded sequence in a cDNA clone. The polypeptide and its polynucleotide
 CC are useful for treating, preventing, diagnosing and/or prognosing
 CC diseases and/or disorders such as pathological cell proliferative
 CC neoplasias e.g. acute myelogenous leukemias, adenocarcinoma; respiratory
 CC disorders such as chronic rhinitis, sinusitis; immunodeficiencies such as
 CC X-linked agammaglobulinemia, X-linked infantile agammaglobulinemia;
 CC inflammatory disorders such as adenitis, alveolitis; immune complex
 CC diseases such as serum sickness, polyarthritis nodosa; bleeding disorders
 CC such as thrombocytopenia, Von Willebrand's disease; acquired platelet
 CC dysfunction such as kidney failure, multiple myeloma; disorders
 CC associated with macrophage numbers and/or macrophage function such as
 CC Gaucher's disease, Niemann-Pick disease, tumours such as colon cancer,
 CC pancreatic cancer; renal disorders such as kidney failure, nephritis;
 CC bone disorders such as Albers-Schonberg disease, bowlegs; muscle
 CC disorders such as Becker's muscular dystrophy, Duchenne's muscular
 CC dystrophy; nervous disorders such as ischaemic lesions, traumatic lesions
 CC; endocrine disorders such as Cushing's syndrome, corticosteroid

Query Match 37.8%; Score 1830; DB 9; Length 1830;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1830; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 201 AGATGAAACAGTTCTTCTACTTTATTGGATACAAAGTGAATTAAGATTCTAGATCCCC 260
 DB 1 AGATGAAACAGTTCTTCTACTTTATTGGATACAAAGTGAATTAAGATTCTAGATCCCC 60
 QY 261 TTCTCACCGGTGTCANTTACCTCTCTTGGCCAGTGTGAATGAATCTCAGTTCTTAA 320
 DB 61 TTCTCACCGGTGTCANTTACCTCTCTTGGCCAGTGTGAATGAATCTCAGTTCTTAA 120
 QY 321 TGAGTCACACCAACCACTGAAAGTCTTCTCCCTGGCTCATTTCAGCTCCCTGACCACAGA 380
 DB 121 TGAGTCACACCAACCACTGAAAGTCTTCTCCCTGGCTCATTTCAGCTCCCTGACCACAGA 180
 QY 381 GGAGAGGATCACTGTGCTAATGACAGGACTGTAATCTAATCCAGAGATTGCCAAT 440
 DB 181 GGAGAGGATCACTGTGCTAATGACAGGACTGTAATCTAATCCAGAGATTGCCAAT 240
 QY 441 GTGGATTGATGAAATGCTGTTCAGAGAGACCACTTAATTAAGAGAACTATAGTTGGGA 500
 DB 241 GTGGATTGATGAAATGCTGTTCAGAGAGACCACTTAATTAAGAGAACTATAGTTGGGA 300
 QY 501 TGATCAATGCAGTGTGTGAGTGGGAGAGAGAAATGTGGAAACCTGGCTTGTCTGCC 560
 DB 301 TGATCAATGCAGTGTGTGAGTGGGAGAGAGAAATGTGGAAACCTGGCTTGTCTGCC 360
 QY 561 AGATGAGAAGAAATGTTCTTGTGTAGCCGTCATGCATAACTGTGATAAAAGGACATTACA 620
 DB 361 AGATGAGAAGAAATGTTCTTGTGTAGCCGTCATGCATAACTGTGATAAAAGGACATTACA 420
 QY 621 AAACGATTTACAGGATTGTAATTAATTAATAGTCAATCCCTTATGATGCTTTAGCTG 680
 DB 421 AAACGATTTACAGGATTGTAATTAATTAATAGTCAATCCCTTATGATGCTTTAGCTG 480
 QY 681 TTCACTGGATAATGAAACACAGACAACTGATCAATTTAGTTTATAGTATAAATGAGTCCAC 740
 DB 481 TTCACTGGATAATGAAACACAGACAACTGATCAATTTAGTTTATAGTATAAATGAGTCCAC 540
 QY 741 TGAAGAGATATGAATTCAGAGAGAAATGGATCCATTGAATAGACCGGAAACAGAGGG 800
 DB 541 TGAAGAGATATGAATTCAGAGAGAAATGGATCCATTGAATAGACCGGAAACAGAGGG 600
 QY 801 GAGATCTGTTAAACCATCTGTGCTCTCTTCTCATCTGATAGTCTAGCCAGTGTCTGTTCCCC 860
 DB 601 GAGATCTGTTAAACCATCTGTGCTCTCTTCTCATCTGATAGTCTAGCCAGTGTCTGTTCCCC 660
 QY 861 TTCACATTAAAGATGACCGAGATAGGTAGAGACCCCTCCATGCTCTCGGATTACAAG 920


```
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 03-SEP-2000; 2000US-0229509P.
PR 03-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 12-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 21-SEP-2000; 2000US-0234937P.
PR 25-SEP-2000; 2000US-0234998P.
PR 25-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241783P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2000US-0254097P.
PR 17-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764878.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Barash SC;
PI WPI; 2003-695900/66.
XX
XX Novel isolated lung antigen polypeptides useful for treating, preventing,
PT diagnosing acute myelogenous leukemias, adenocarcinoma, thrombocytopenia,
PT Von Willebrand's disease.
XX
XX Disclosure; SEQ ID NO 356; 178pp; English.
XX
XX The invention relates to an isolated lung antigen polypeptide sequence or
CC encoded sequence in a cDNA clone. The polypeptide and its polynucleotide
CC are useful for treating, preventing, diagnosing and/or prognosing
CC diseases and/or disorders such as pathological cell proliferative
CC neoplasias e.g. acute myelogenous leukemias, adenocarcinoma; respiratory
CC disorders such as chronic rhinitis, sinusitis; immunodeficiencies such as
CC X-linked agammaglobulinaemia, X-linked infantile agammaglobulinaemia;
CC inflammatory disorders such as adenitis, alveolitis, immune complex
CC diseases such as serum sickness, polyarteritis nodosa; bleeding disorders
CC such as thrombocytopenia, Von Willebrand's disease; acquired platelet
CC dysfunction such as kidney failure, multiple myeloma; disorders
CC associated with macrophage numbers and/or macrophage function such as
CC Gaucher's disease, Niemann-Pick disease; tumours such as colon cancer,
CC pancreatic cancer; renal disorders such as kidney failure, nephritis;
CC bone disorders such as Albers-Schönberg disease; bowlegs; muscle
CC disorders such as Becker's muscular dystrophy, Duchenne's muscular
CC dystrophy; nervous disorders such as ischaemic lesions, traumatic lesions
CC ; endocrine disorders such as Cushing's syndrome, corticosteroid
Query Match 37.8%; Score 1830; DB 9; Length 1830;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1830; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 201 AGATGAACAGTTCTTCTACTTTATGGATACAAAGTGGATGAATAGATTCTAGATCCCC 260
D5 1 AGATGAACAGTTCTTCTACTTTATGGATACAAAGTGGATGAATAGATTCTAGATCCCC 60
QY 261 TTCTACCGGCTGTCATTTAAACCTACTTTGGCCAGTGTGAATGAATCTGCAGTTTCTAA 320
```


XX 03-FEB-2000.
 PD 20-JUL-1999; 99WO-CA000656.
 PF 20-JUL-1998; 98CA-02237701.
 PR 10-DEC-1998; 98CA-02233647.
 XX (HSCR-) HSC RES & DEV LP.
 PA Wzana JL;
 XX WPI; 2000-182691/16.
 XX P-PSDB; AAV44751.
 DR
 XX New Smad Anchor for Receptor Activation (SARA) polypeptides, useful for
 PT developing agents for treating e.g. wounds, arthritis, fibrosis,
 PT inflammation, tumors, fibrogenesis or tissue damage.
 XX
 PS Claim 9; Page 57-58; 93pp; English.
 XX
 CC The present sequence is the cDNA encoding the Xenopus Smad Anchor for
 CC Receptor Activation protein-1 (XSARAL), a TGF-beta modulator. It is
 CC isolated from lambda ZAP II xenopus dorsal lip library and has 62%
 CC identity to hSARAL. The XSARA proteins bind to receptor-regulated Smad
 CC proteins and ensures its appropriate localisation for activation by a
 CC type I receptor of a TGF-beta, activin or bone morphogenetic protein
 CC (BMP) signalling pathway. SARA proteins has anti-inflammatory,
 CC cytostatic, anti-arthritis and vulnary activity. The SARA proteins are
 CC useful for developing agents for prevention or treatment of disorders
 CC involving TGF-beta superfamily member signalling pathway. Such agents can
 CC be used for wound healing, scarring, arthritis and fibrosis (e.g. liver
 CC and kidney), in modulating inflammatory and immune responses, tumour
 CC progression, cell proliferation or fibrogenesis and in tissue
 CC morphogenesis. It is also used for protecting, restoring and regenerating
 CC tissues after tissue damage
 XX
 SQ Sequence 4823 BP; 1405 A; 1035 C; 1089 G; 1294 T; 0 U; 0 Other;
 Query Match 31.9%; Score 1545.8; DB 3; Length 4823;
 Best Local Similarity 71.8%; Pred. No. 0;
 Matches 2136; Conservative 0; Mismatches 792; Indels 45; Gaps 7;
 QY 1443 ACAGTGTGTGGATTGGCAGATGCGAGTCTAGATTAAAGCACTTGCATTAGTGAAG 1502
 DB 1089 AGAGTTTGAAGAGGGGTAGATGCTTAATGCTTGTGGAATAATGATGTGTAGCTTATA 1148
 QY 1503 TGAAGAATGTGATTTCTCCACTGTTATAGACACACAGCAGCAAAATTTCTATCTAATGG 1562
 DB 1149 AGAAGCAATAGATTTGGCTGAAGAAATGGAACATAATGCACCAATGCTCTGTACAATGG 1208
 QY 1563 TTGTGATCTCTATGGAATGCAAGCCAGGTGTTCTTTTCTCCAAAGCTTTACCTC 1622
 DB 1209 GHTGATTCCTATGGAATGAAMAAACCCAGCGTAGCTCAAAACCAAGAAATTTACCTTC 1268
 QY 1623 CAAGAAGATTCAGTAACAGAGAAAGAAATAAGAGAAAGCAAGTCAAGTCTACTC 1682
 DB 1269 AAAAGAAGATTCGTGACAGAGAAAGAAATAAGAGAAAGCAAGTCAAGTCTACTC 1328
 QY 1683 ABATATTTATG---RACAGAGAGGAATGCGGCACACAGAGGAGTGGACTCTTTTAA 1739
 DB 1329 TGGTGTATTGAACACAAAGAGAGATGATGTTTACAGAGAGAGTGGACTCTCTGTATA 1388
 QY 1740 CAGCACTGGTGAATTAATGAAGAAATTTATTTACATAATTTCTGTAGTCAAGTTCATC 1799
 DB 1389 TGCCTAAGGCTGACCAATGAAGCA---ATTGCTAGTCTTTGTAATCAGGTTCCATC 1445
 QY 1800 AGTGTCTGGGCAATCTTCCCAAGTAGTAGCAGCTGCACTATCATGTTCTCTTT 1859
 DB 1446 CATGATGGGCAACATCACCACAAAGAGGCAAGTGTGCAATCTCTCAGTGTTCATA 1505
 QY 1850 TGGTGGTGCAAGACCAAGCAACCTTCTAATCTTAAACTTCAATTCCAAAGCCATTATC 1919

DB 1506 CGGTGAGCAGCAGCACTAAGCAGCACTCATCTCAAACTCCATATTTCCAAAGCCATTGTC 1565
 QY 1920 AGACCATTTACAAA---ATGACTTTCTCTGCAAAAGAGTGAATAATACTATAAAATAAAAA 1976
 DB 1566 TGAATGTTGAGAGCGATCTCAATCTCCCAATGCTGGTGCAGCTCTTAANTACAAAA 1625
 QY 1977 TGATATTTCTGGGAAGCAAAATTAGGGGAAATCTAGCAACCAATGTATGATGATCCATC 2036
 DB 1626 TGACATGTTAAACAAATCAAAATCAGGGGGATAAAGCTGATTTTCAAGATCACTCGCTGAGGA 1685
 QY 2037 TTTGGAAACATCTCTAATGTGATACAAATAGGGGAAACATTTAGAAAGTTATAGAGGTGA 2096
 DB 1686 TTCTGACGTGCGCAGCCCTGT---TACTGATGCTAATGGTGATTTCCCTGGAGATACAG 1742
 QY 2097 GATCTCCACTAGACCATGCTTGTGATTTAGCTCCAGATAGCCAGATTAATGATCTCAGAGC 2156
 DB 1743 GGGACCTGGCAGCTTTGCTCTTCCAGAGAGCCAGACCAACCAATCTGCTGCTTC 1802
 QY 2157 TGGTCAGTTTGGAAATTTCTGCGCAGAAAGCCATTCAACACGCTGGGTGAGGGTCTCCAGT 2216
 DB 1803 CGGGCAGTTGGGTACCCATCTCTAAGCCATTTACTACTAGGGAAGTGGCTCCAGT 1862
 QY 2217 ATGGGTACGGGATTTCTAGGCTCCAAATTTGATGAATGTGAAGCCAGGTTTACATTCAC 2276
 DB 1863 CTGGGTGCGGATTTCCCAAGCACCACCAACTGATGAAGTGCAGGCCAGATTTTACATTTAC 1922
 QY 2277 CAAAAGGAGGCATCACTGCAGAGCATGTGGGAAGGTTTCTGTCTCTCTCTCTAGCCT 2336
 DB 1923 CAAAAGGAGGCATCACTGCCAGCTTTGGGAAGAGTGTCTGTCTCTCTCTGTCTGCTCT 1982
 QY 2337 GAAATGTAACCTGTTATACATGTGACAGAAAGAGCTAGAGTGTGTATCTTGCCTATTC 2396
 DB 1983 AAAATGCCAACTACAGTACATGAGTAAAAAGAGGCTCGTGTGTGTATTTGCTCATTC 2042
 QY 2397 AGTGTATATGATGCTCAAGCTCGGGAAGACATCATGATGCTCCCAAGCAGAGCCCTAA 2456
 DB 2043 TGTGCTTATGATGCTCAAGCATGGGAGAACATTTAAGTGCATCGTCCAAAGCCCAA 2102
 QY 2457 CCTAACCAATCTCTGTAATCTGTTCTTACTATCCCTCCCTTGCAGCAAGCTCAGGCTC 2516
 DB 2103 TCCAAATATCTGCTGTAATCTGCTCAACTATCCCTCCGATGAGCAGGCAAGCTTC 2162
 QY 2517 AGGAGCTCTGAGCTCTCCACCTCCACCTGCTGATGCTGAGTGTGGAGTGTAAAGACACC 2576
 DB 2163 AGGAGCATGTAGTTCCTCCACCTCCACCTGCTGATGCTGAGTGTGGTGTAAACATCC 2222
 QY 2577 TGGACAGAAAGTGGCTCAGCCAGAGAGCAGAGCGAGTTTGGTTTCTGATGGATCTT 2636
 DB 2223 AGGAACTGAAGGTCACAGTCAAGGAAACAGCGCGGTGTTTGGTTTCTGATGGAATAT 2282
 QY 2637 GCCCAATGAGAGTGTGCTGATGAGCCAAATTAACATGAATGGAATCTTCTCTGAGG 2696
 DB 2283 ACCCAAGGAGAGACTGCTGACTCAGATAATGCA-----AACGTAACACTACAGTGGCTGG 2336
 QY 2697 AACCTGCTGTGTACACGACCCAGCTCAAGCCAGTAACTACCACTCTCTACCAAGCAGA 2756
 DB 2337 GACACTTACTGTGTCACATACC-----AACAAATTCACATCTTTCAGAGTC 2381
 QY 2757 GACGATATTTGCTTATTTCTGCGGAGTATAACTCAGTGGAGTCTCTGTGAGTGC 2816
 DB 2382 TGAGAACCTTCTGGAATTTCTGGAAGTATAACTCAG-----GTTGCGAGTGC 2429
 QY 2817 AATGAATCTTATCTCTGAAGTGGCTTCTCTCCCATTTCTCATCTCCACTCGTGTAAAGG 2876
 DB 2430 AATGAACCTTATTCAGAAAGATGGCTTCTCTCTTACTAATCTCTACTCGAGTAAAAAGG 2489
 QY 2877 AGACTATGCTGTGAAGAGAAACCATCAGATTTCAATTCAGTAACTCAGAGTGGAGGATGG 2936
 DB 2490 AGATTACCGATTTGAGGAACGCCCTTCCAGATGCTGTGATGAGCAACTAGAGGAAGG 2549
 QY 2937 TGGCCCTCAGCCACTGTGTATTTGTTTAAATGCAAAATTTGTGTCAATGGTGTAAATTTGT 2996
 DB 2550 AGGACCAATCCTTTGGTTTTGTTTCTTAAATGCAAACTCTTTGGCCCATGTTTGAATCGT 2609

sequences given in the specification (including its fragment or variant), a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences given in the specification (including its fragment or variant), a vector comprising the SID (RTM) polynucleotide, a recombinant host cell comprising the vector, a protein chip comprising the polypeptides and a record comprising all or part of the data, listed in the specification. The complex, polypeptides, polynucleotides and compounds are useful for preventing or treating metabolic disorders such as obesity or diabetes. The polynucleotides are useful as probes or primers. The complex is particularly useful for identifying selected interacting domains (SID (RTM)) for screening drugs that modulate the protein interaction, thus exhibiting the therapeutic effect. The present sequence encodes a SID (prey) protein of the invention

Sequence 888 BP; 248 A; 180 C; 224 G; 236 T; 0 U; 0 Other;

Query Match	14.5%;	Score	699.4;	DB	7;	Length	888;
Best Local Similarity	83.3%;	Pred.	No. 7.8e-182;				
Matches	887;	Conservative	0;	Mismatches	1;	Indels	177;
Gaps							

QY	2003	GGGAAACTCAGCAACCAATGTATGCAGTCCCATCTTTGGGAACAATCTCTAATGTCGATA	2062
Db	1	GGGAAACTCAGCAACCAATGTATGCAGTCCCATCTTTGGGAACAATCTCTAATGTCGATA	60
QY	2063	CAAATGGGAAACATTTAGAAAAGTTATGAGCGTGAGATCTCCACTAGACCATGCCCTGGAT	2122
Db	61	CAAATGGGAAACATTTAGAAAAGTTATGAGCGTGAGATCTCCACTAGACCATGCCCTGGAT	120
QY	2123	TAGCTCCAGATAGCCAGCAATAATGATCTCAGAGCTGGTCAGTTTGGAAATTTCTGCGAGAA	2182
Db	121	TAGCTCCAGATAGCCAGCAATAATGATCTCAGAGCTGGTCAGTTTGGAAATTTCTGCGAGAA	180
QY	2183	AGCCATTACCACGCTGGGTGAGTGGCTCCAGTATGGGTACCGGATTTCTCAGGCTCCAA	2242
Db	181	AGCCATTACCACCTCTGGGTGAGTGGCTCCAGTATGGGTACCGGATTTCTCAGGCTCCAA	240
QY	2243	ATTGCATGAATGTGAAGCCAGCTTTACATTCACCAAAAGGAGGCATCACTGCAGAGCAT	2302
Db	241	ATTGCATGAATGTGAAGCCAGCTTTACATTCACCAAAAGGAGGCATCACTGCAGAGCAT	300
QY	2303	GTGGGAAGTTTTCTGTGCTTCCTGTGTAGCCTGAAATGTAACCTGTTATCATGGACA	2362
Db	301	GTGGGAAGTTTTCTGTGCTTCCTGTGTAGCCTGAAATGTAACCTGTTATCATGGACA	360
QY	2363	GAAGAAGCTAGAGTGTGTAACTCTGCCATTCAGTCTAAAGTGAAGTCAAGCCTGGG	2422
Db	361	GAAGAAGCTAGAGTGTGTAACTCTGCCATTCAGTCTAAAGTGAAGTCAAGCCTGGG	407
QY	2423	AGACATGATGAGTGGCTCAAGCCAGAGCCCTAACCCCTAACCAATCTGCTGAATACTGTT	2482
Db	408	-----	407
QY	2483	CTACTATCCCTCCCTTGAGCAAGCTCAGGCCTCAGAGCTCTGAGCTCTCCACCTCCCA	2542
Db	408	-----	407
QY	2543	CTGTGATGGTACCTGTGGAGTTTTAAAGCACCCCTGGAGCAGAAGTGGCTCAGCCCCAGAG	2602
Db	408	-----GTGGCTCAGCCCCAGAG	423
QY	2603	AGCAGAGCGAGTTTGGTTTGTCTGATGGGATCTTGCCCAATPGAGAGAAGTTGCTGATGCAG	2662
Db	424	AGCAGAGCGAGTTTGGTTTGTCTGATGGGATCTTGCCCAATPGAGAGAAGTTGCTGATGCAG	483
QY	2663	CCAATTTAAACATGAAATGGAATCTTCTGAGAAACCTGAGAAACCTGGCTGTCAACAGACCCAG	2722
Db	484	CCAATTTAAACATGAAATGGAATCTTCTGAGAAACCTGAGAAACCTGGCTGTCAACAGACCCAG	543
QY	2723	TCAAGCCAGTAACCTACCAAGTCTCTTACCAGCAGAGAAGGATATTGTTCTATTCTCTGGGA	2782
Db	544	TCAAGCCAGTAACCTACCAAGTCTCTTACCAGCAGAGAAGGATATTGTTCTATTCTCTGGGA	603
QY	2783	GTATAACTCAGGTTGGAAAGTCTGTTTGGAAATGMAATGAAATCTTATTCTCTGAAGATGCC	2842

Db	604	GTATAACTCAGGTTGGAAGTCCTGTTGGAGTGCAATGAATCTATTCTCGAAGATGGCC	663
Qy	2843	TTTCCTCCCAATTCATCTCCACTGGGTAAAGGAGACTATGCTGTGGAAGAGAAACCAT	2902
Db	664	TTTCCTCCCAATTCATCTCCACTGGGTAAAGGAGACTATGCTGTGGAAGAGAAACCAT	723
Qy	2903	CACAGATTTTCAGTAATGCGACAGTTGAGAGATGGTGGCCCTGACCCACITGTATTGTTT	2962
Db	724	CACAGATTTTCAGTAATGCGACAGTTGAGAGATGGTGGCCCTGACCCACITGTATTGTTT	783
Qy	2963	TAAATCGAATTTGTTGTCAATGGTTTAAATTTCTAAATATGTGAACAGGAAGTCTGGT	3022
Db	784	TAAATCGAATTTGTTGTCAATGGTTTAAATTTCTAAATATGTGAACAGGAAGTCTGGT	843
Qy	3023	GTTTTCAACCAAGGGAATGATGCGAGTGGTCACTGAGATAG	3067
Db	844	GTTTTCAACCAAGGGAATGATGCGAGTGGTCACTGAGATAG	888
RESULT 9			
ACAS7333			
ID ID ACAS7333 standard; cDNA; 576 BP.			
XX	ACAS7333;		
XX	10-JUN-2003 (first entry)		
XX	Human adipocyte Selected Interacting domain, SID, cDNA #420.		
XX	Human; ss; gene; prey; adipocyte; SID; selected interacting domain;		
XX	anorectic; antidiabetic; protein-protein interaction; diabetes;		
XX	yeast 2-hybrid assay; metabolic disorder; obesity.		
XX	Homo sapiens.		
XX	WO200286122-A2.		
XX	31-OCT-2002.		
XX	14-MAR-2002; 2002WO-EP003768.		
XX	14-MAR-2001; 2001US-0275734P.		
XX	(HYBR-) HYBRIGENICS.		
XX	Legrain P, Daviet L;		
XX	WPI; 2003-103412/09.		
XX	P-PSDB; ABU070789.		
XX	New complex between two interacting proteins in adipocyte cells, useful for identifying selected interacting domains that modulate protein interactions, or for preventing or treating metabolic disorders such as obesity or diabetes.		
XX	Claim 7; Page 250; 382pp; English.		
XX	The invention relates to a complex between two interacting proteins in adipocyte cells, given in the specification. The proteins are identified by selecting a bait protein from a known adipocyte marker and then performing a yeast 2-hybrid selection to isolate prey proteins encoded by members of an adipocyte cDNA library. The proteins are designated SID (RTM) (selected interacting domains) proteins. Also included are a polynucleotide encoding a polypeptide in the adipocyte cells, a recombinant host cell expressing at least one of the interacting polypeptides of the complex, selecting a modulating compound in adipocyte cells, a SID (RTM) polypeptide comprising any of the 738 amino acid sequences given in the specification (including its fragment or variant), a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences given in the specification (including its fragment or variant), a vector comprising the SID (RTM) polynucleotide, a recombinant host cell comprising the vector, a protein chip comprising the polypeptides and a		

CC record comprising all or part of the data, listed in the specification.
 CC The complex, polypeptides, polynucleotides and compounds are useful for
 CC preventing or treating metabolic disorders such as obesity or diabetes.
 CC The polynucleotides are useful as probes or primers. The complex is
 CC particularly useful for identifying selected interacting domains (SID
 CC (RTM)) for screening drugs that modulate the protein interaction, thus
 CC exhibiting the therapeutic effect. The present sequence encodes a SID
 CC (prey) protein of the invention

XX Sequence 576 BP; 149 A; 134 C; 138 G; 155 T; 0 U; 0 Other;

Query Match 11.9%; Score 576; DB 7; Length 576;

Best Local Similarity 100.0%; Pred. No. 6.8e-148;

Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2457 CCTTAAATCTCTGCTGAATCTCTTCTACTATCCCTCCCTTGCAGCAAGCTCAGGCCTC 2516

DB 1 CCTTAAATCTCTGCTGAATCTCTTCTACTATCCCTCCCTTGCAGCAAGCTCAGGCCTC 60

QY 2517 AGGAGCTCTGAGCTCTCCACCTCCACCTGCTGATGCTGAGTGGAGTTTAAACACCC 2576

DB 61 AGGAGCTCTGAGCTCTCCACCTCCACCTGCTGATGCTGAGTGGAGTTTAAACACCC 120

QY 2577 TGGAGCAGAGTGGCTCTCAGCCAGCAGAGCAGAGCGAGTTTGGTTTGTCTGATGGATCTT 2636

DB 121 TGGAGCAGAGTGGCTCTCAGCCAGCAGAGCAGAGCGAGTTTGGTTTGTCTGATGGATCTT 180

QY 2637 GCCCAATGAGAGTGGCTGATGCGACCAATTAACATTAATGAATCTCTCTGCGAGG 2696

DB 181 GCCCAATGAGAGTGGCTGATGCGACCAATTAACATTAATGAATCTCTCTGCGAGG 240

QY 2697 AACCTTGGCTGTGTCTACACAGCAGCCAGTCAAGCCAGTAACTACCACTCTCTACCAAGCAGA 2756

DB 241 AACCTTGGCTGTGTCTACACAGCAGCCAGTCAAGCCAGTAACTACCACTCTCTACCAAGCAGA 300

QY 2757 GACGATATTTGCTATCTCTGGAGTATCACTCAGTGGAGTCTCTGTTGGAAGTGC 2816

DB 301 GACGATATTTGCTATCTCTGGAGTATCACTCAGTGGAGTCTCTGTTGGAAGTGC 360

QY 2817 AATGAATCTTATCTCTGAAGATGGCTTCCCTCCATCTCTCCACTCTCTGTTGGAAGG 2876

DB 361 AATGAATCTTATCTCTGAAGATGGCTTCCCTCCATCTCTCCACTCTCTGTTGGAAGG 420

QY 2877 AGACTATCTCTGGAAGAGAAACCATCAAGATTCAGTATGACGAGTTGGAGGATGG 2936

DB 421 AGACTATCTCTGGAAGAGAAACCATCAAGATTCAGTATGACGAGTTGGAGGATGG 480

QY 2937 TGGCCCTGACCCACTTGTATTTGTTTAAATGCAAAATTTGTTCAATGTTAAATTTGT 2996

DB 481 TGGCCCTGACCCACTTGTATTTGTTTAAATGCAAAATTTGTTCAATGTTAAATTTGT 540

QY 2997 AAATTATGTAACAGGAGTCTGTTGTTTCAACAC 3032

DB 541 AAATTATGTAACAGGAGTCTGTTGTTTCAACAC 576

RESULT 10

AAS29903

ID AAS29903 standard; cDNA; 610 BP.

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

fungus infection; viral infection; ocular disorder; endocrine disorder;
 gastrointestinal disorder; renal disorder; respiratory disorder;
 wound healing; skin aging; organ transplantation; food preservative;
 tissue regeneration; anti-infertility; food additive.

Homo sapiens.

XX WO200155303-A2.

XX PD 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001301.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 11-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0224519P.

XX 14-AUG-2000; 2000US-0225213P.

XX 14-AUG-2000; 2000US-0225214P.

XX 14-AUG-2000; 2000US-0225266P.

XX 14-AUG-2000; 2000US-0225267P.

XX 14-AUG-2000; 2000US-0225268P.

XX 14-AUG-2000; 2000US-0225270P.

XX 14-AUG-2000; 2000US-0225447P.

XX 14-AUG-2000; 2000US-0225757P.

XX 14-AUG-2000; 2000US-0225758P.

XX 18-AUG-2000; 2000US-0225759P.

XX 22-AUG-2000; 2000US-0226279P.

XX 22-AUG-2000; 2000US-0226681P.

XX 22-AUG-2000; 2000US-0226868P.

XX 22-AUG-2000; 2000US-0227182P.

XX 23-AUG-2000; 2000US-0227009P.

XX 30-AUG-2000; 2000US-0228924P.

XX 01-SEP-2000; 2000US-0229287P.

XX 01-SEP-2000; 2000US-0229343P.

XX 01-SEP-2000; 2000US-0229344P.

XX 05-SEP-2000; 2000US-0229345P.

XX 05-SEP-2000; 2000US-0229505P.

XX 06-SEP-2000; 2000US-0229513P.

XX 06-SEP-2000; 2000US-0230437P.

XX 08-SEP-2000; 2000US-0230438P.

XX 08-SEP-2000; 2000US-0231242P.

XX 08-SEP-2000; 2000US-0231243P.

XX 08-SEP-2000; 2000US-0231244P.

XX 08-SEP-2000; 2000US-0231413P.

XX 08-SEP-2000; 2000US-0231414P.

XX 08-SEP-2000; 2000US-0232080P.

XX 08-SEP-2000; 2000US-0232081P.

XX 12-SEP-2000; 2000US-0231968P.

XX 14-SEP-2000; 2000US-0232397P.

XX 14-SEP-2000; 2000US-0232398P.

XX 14-SEP-2000; 2000US-0232399P.

XX 14-SEP-2000; 2000US-0232400P.

XX 14-SEP-2000; 2000US-0232401P.

XX 14-SEP-2000; 2000US-0233063P.

XX 14-SEP-2000; 2000US-0233064P.

XX 14-SEP-2000; 2000US-0233065P.

PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 26-SEP-2000; 2000US-0235485P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 28-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250319P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251858P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.

PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457723/49.
XX P-PSDB; AAU18616.
XX Isolated polypeptide for treating, preventing and/ or prognosing
PT respiratory disorders related to the lung including lung cancers and also
PT for testing and detection e.g. diagnosis.
XX
XX
PS Claim 1; SEQ ID NO 75; 507pp; English.

CC Sequences AAS29839-AAS29930 represent cDNA molecules, which encode the
CC lung antigen polypeptides of the invention. Lung antigen polypeptides and
CC their associated polynucleotides are useful in the diagnosis, treatment
CC and prevention of various types of disorders in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. A pathological
CC condition can be determined by detecting the presence or absence of a
CC mutation in a lung antigen polynucleotide. The treatable disorders
CC include autoimmune diseases such as rheumatoid arthritis.
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma and pleurisy. The polypeptides can
CC also be used to aid wound healing, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

Query Match 11.8%; Score 570.4; DB 5; Length 610;

Best Local Similarity 99.2%; Pred. No. 2.5e-146;

Matches 593; Conservative 1; Mismatches 2; Indels 2; Gaps 2;

QY 1 GCNACTGATCAGCAGGACTGGCTGGTGCAGCAGACATCATGAGTAAGCAGCGA 60
DB 1 GCNACTGATCAGCAGGACTGGCTGGTGCAGCAGACATCATGAGTAAGCAGCGA 60
QY 61 AGTCTCTTCTTATCATCGTGTGTAAGGGGAAAAGGTTTAAACAAGTCTCTTAAGTGGTG 120
DB 61 AGTCTCTTCTTATCATCGTGTGTAAGGGGAAAAGGTTTAAACAAGTCTCTTAAGTGGTG 120
QY 121 TTTCTCCTCAGGATGAGAAATTACTTCAAGCAGAGACTTACACCTGGCAGAGGTGTTA 180
DB 121 TTTCTCCTCAGGATGAGAAATTACTTCAAGCAGAGACTTACACCTGGCAGAGGTGTTA 180
QY 181 GATGAATTTGAACAAAACGAAGATGAAACAGATTTCTTCTACTTTATTGGATACAAAGTGG 240
DB 181 GATGAATTTGAACAAAACGAAGATGAAACAGATTTCTTCTACTTTATTGGATACAAAGTGG 240
QY 241 AATAAGATTTAGATCCCTCTCTACCGGTGTGTATTTAACCTACTTTGGCAGTGTG 300
DB 241 AATAAGATTTAGATCCCTCTCTACCGGTGTGTATTTAACCTACTTTGGCAGTGTG 300
QY 301 AATGAATCTGCAGTTTCTAATGAGTCACAAACCAACTGAAAGTCTTCTCCCTGGTCTAT 360
DB 301 AATGAATCTGCAGTTTCTAATGAGTCACAAACCAACTGAAAGTCTTCTCCCTGGTCTAT 360
QY 361 TCAGCTCCCTGACACAGAGGAGGATCACTGTGCTAATGAGCAGAGCTGTATCTA 420
DB 361 TCAGCTCCCTGACACAGAGGAGGATCACTGTGCTAATGAGCAGAGCTGTATCTA 420
QY 421 AATCCAGAGATTCCCAATGATGGATTGATGAAATGCTGTTTCAGAGACCAAGTTAATT 480

Db 421 AATCCAGAGATGCCCAATGTCGATTCGATGAAATGCTGTCAGAAAGACCAAGTAAAT 480
QY 481 AAGAGAACTAGTGGATGATCAATCCAGTCTGTTGAAGTGGGAGAGAAATGT 540
Db 481 AAGAGAACTAGTGGATGATCAATCCAGTCTGTTGAAGTGGGAGAGAAATGT 540
QY 541 GGAACCTGGCTTGTCTGCAGATGAGAAGATGTTCTGTTGAGCCGTCATGCATA 598
Db 541 GG-AACTGGCTTGTCTGCAGATGAGAAGATGTTCTGTTGAGCCGTCATGCATA 596

RESULT 11
ADB33148
ID ADB33148 standard; cdna; 610 BP.
XX
AC ADB33148;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human novel lung related polypeptide cdna SEQ ID NO 75.
KW gene therapy; lung antigen; neoplasia; acute myelogenous leukaemia;
KW adenocarcinoma; respiratory disorder; chronic rhinitis; sinusitis;
KW immunodeficiency; X-linked agammaglobulinaemia;
KW X-linked infantile agammaglobulinaemia; inflammatory disorder;
KW adrenailitis; alveolitis; immune complex disease; serum sickness;
KW polyarteritis nodosa; bleeding disorder; thrombocytopenia;
KW Von Willebrand's disease; acquired platelet dysfunction; kidney failure;
KW multiple myeloma; macrophage related disorder; Gaucher's disease;
KW Neimann-Pick disease; tumour; colon cancer; pancreatic cancer;
KW renal disorder; nephritis; bone disorder; Albers-Schonberg disease;
KW bowleg; muscle disorder; Becker's muscular dystrophy;
KW Duchenne's muscular dystrophy; nervous disorder; ischaemic lesion;
KW traumatic lesion; endocrine disorder; Cushing's syndrome;
KW corticosteroid deficiency; gastrointestinal disorder; dysphagia;
KW gastric reflux; human; ss; gene.
XX
OS Homo sapiens.
XX
PN US2003054368-A1.
XX
PD 20-MAR-2003.
XX
PF 22-FEB-2002; 2002US-00079854.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 05-SEP-2000; 2000US-0230437P.
PR 05-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0231977P.
PR 14-SEP-2000; 2000US-0231988P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 01-NOV-2000; 2000US-0244618P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.

PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 06-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 PR 17-JAN-2001; 2001US-00764878.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Ruben SM, Barash SC;
 WPI; 2003-695900/66.
 P-PSDB; ADB33240.
 Novel isolated lung antigen polypeptides useful for treating, preventing,
 diagnosing acute myelogenous leukemias, adenocarcinoma, thrombocytopenia,
 Von Willebrand's disease.
 Claim 3; SEQ ID NO 75; 178pp; English.
 The invention relates to an isolated lung antigen polypeptide sequence or
 encoded sequence in a cDNA clone. The polypeptide and its polynucleotide
 are useful for treating, preventing, diagnosing and/or prognosing
 diseases and/or disorders such as pathological cell proliferative
 neoplasias e.g. acute myelogenous leukemias, adenocarcinoma; respiratory
 disorders such as chronic rhinitis, sinusitis; immunodeficiencies such as
 X-linked agammaglobulinemia, X-linked infantile agammaglobulinemia;
 inflammatory disorders such as adenitis, alveolitis; immune complex
 diseases such as serum sickness, polyarteritis nodosa; bleeding disorders
 such as thrombocytopenia, Von Willebrand's disease; acquired platelet
 dysfunction such as kidney failure, multiple myeloma; disorders
 associated with macrophage numbers and/or macrophage function such as
 Gaucher's disease, Niemann-Pick disease; tumours such as colon cancer,
 pancreatic cancer; renal disorders such as kidney failure, nephritis;
 bone disorders such as Albers-Schonberg disease, bowlegs; muscle
 disorders such as Becker's muscular dystrophy, Duchenne's muscular
 dystrophy; nervous disorders such as ischaemic lesions, traumatic lesions
 Query Match 11.8%; Score 570.4; DB 9; Length 610;
 Best Local Similarity 99.2%; Pred. No. 2.5e-146;
 Matches 593; Conservative 1; Mismatches 2; Indels 2; Gaps 2;

QY 1 GCATACGAATCAGCAGGACTGGCTGGTGGTGACGACAGATCATGATAGTAAACCCGAGA 60
 DB |||||
 1 GCATACGAATCAGCAGGACTGGCTGGTGGTGACGACAGATCATGATAGTAAACCCGAGA 60
 QY 61 AGTCTGTTCCCTTATCAGCTGTGTAAAGGGGAAAAAGGTTTAAACAAGTCTCTTAAGTGGTG 120
 DB |||||
 61 AGTCTGTTCCCTTATCAGCTGTGTAAAGGGGAAAAAGGTTTAAACAAGTCTCTTAAGTGGTG 120
 QY 121 TTCTCTACCGATGGAGNATTACTTCCAAAGCAGAGCTTCAACCTGGGCAAGGTGTTA 180
 DB |||||
 121 TTCTCTACCGATGGAGNATTACTTCCAAAGCAGAGCTTCAACCTGGGCAAGGTGTTA 180
 QY 181 GATGAATTTGAACAAAACGAAGATCAACACAGTCTTCTTCTACTTTATTGGATCAAAAGTGG 240
 DB |||||
 181 GATGAATTTGAACAAAACGAAGATCAACACAGTCTTCTTCTACTTTATTGGATCAAAAGTGG 240
 QY 241 AATAAGATTTAGATCCCCCTTCTCACCGGCTGTCAITTAACCTTACTTTGGCCAGTGTG 300
 DB |||||
 241 AATAAGATTTAGATCCCCCTTCTCACCGGCTGTCAITTAACCTTACTTTGGCCAGTGTG 300
 QY 301 AATGAATCTGCAGTTTCTAAATGAGTCACAACCAACCTGAAAGTCTTCTCCCTGGCTCAT 360
 DB |||||
 301 AATGAATCTGCAGTTTCTAAATGAGTCACAACCAACCTGAAAGTCTTCTCCCTGGCTCAT 360
 QY 361 TCAGTCCCTGACACAGAGAGAGGATCACTGTGCTAATGGACAGACTGTAACTA 420
 DB |||||
 361 TCAGTCCCTGACACAGAGAGAGGATCACTGTGCTAATGGACAGACTGTAACTA 420
 QY 421 AATCCAGAGATTGCCACCAATGTGGATTGATGAAATGCTTTGGCAGAGACCAAGTAAAT 480
 DB |||||
 421 AATCCAGAGATTGCCACCAATGTGGATTGATGAAATGCTTTGGCAGAGACCAAGTAAAT 480
 QY 481 AAGAGAAACTATAGTTGGGATGATCAATCCAGTGTGTGAAGTGGGAGAGAAATGT 540
 DB |||||
 481 AAGAGAAACTATAGTTGGGATGATCAATCCAGTGTGTGAAGTGGGAGAGAAATGT 540
 QY 541 GGAAACCTGGCTTCTGCCAGATGAGAGATGTTCTTGTAGCCGCTCATGATA 598
 DB |||||
 541 GG-AAACCTGGCTTGTGCCAGATGAGAGAAAG-TCCTGRTGTAGCCCTCATGATA 596

RESULT 12

AAZ50069 standard; cDNA; 6632 BP.

AAZ50069;

04-MAY-2000 (first entry)

Human Smad Anchor for Receptor Activation protein-2 encoding cDNA.

Smad Anchor for Receptor Activation protein; hSARA2; human;
 transforming growth factor-beta; TGF-beta; bone morphogenetic protein;
 BMP; activin; anti-inflammatory; cytotatic; antiarthritic; vulnerary;
 TGF-beta modulator; wound healing; scarring; arthritis; immune response;
 inflammatory response; tumour progression; cell proliferation; fibrosis;
 fibrogenesis; tissue morphogenesis; tissue damage; ss.

Homo sapiens.

Key Location/Qualifiers

CDS 249..4868

/tag= a

/product= "Human SARA-2 protein"

/note= "Binds to receptor regulated Smad proteins"

/transl_except= (pos:336..338, aa:Gln)

/transl_except= (pos:822..824, aa:Thr)

W0200005360-A1.

Query Match 11.8%; Score 570.4; DB 9; Length 610;

Best Local Similarity 99.2%; Pred. No. 2.5e-146;

Matches 593; Conservative 1; Mismatches 2; Indels 2; Gaps 2;

20-JUL-1999; 99WO-CA000656.

QY	4441	ACTGTT	4447	
Db	4899	CCTAATT	4905	
RESULT 13				
ID	ABK11709	ABK11709 standard; cDNA; 5402 BP.		
XX	AC	ABK11709;		
XX	DT	05-JUN-2002	(first entry)	
XX	DE	DNA encoding novel human secreted protein #1.		
XX	XX	Secreted protein; gene therapy; immunostimulant; human; stroke;		
KW	KW	immune system disorder; immunologic deficiency syndrome; anaemia;		
KW	KW	ataxia telangiectasia; HIV; human immunodeficiency virus; heart attack;		
KW	KW	Wiskott-Aldrich disorder; thrombocytopenia; scarring; erythematosis;		
KW	KW	autoimmune disease; Addison's disease; encephalomyelitis; lupus;		
KW	KW	multiple sclerosis; autoimmune thyroiditis; diabetes; inflammation;		
KW	KW	nephritis; ischaemia; hypergammaglobulinaemia; Crohn's disease;		
KW	KW	sarcoidosis; Gaucher's disease; cardiovascular disease; telangiectasia;		
KW	KW	persistent truncus arteriosus; arrhythmias; angina pectoris; tumour;		
KW	KW	myocardial infarction; hypertension; neovascularisation; angiogenesis;		
KW	KW	cancer; ocular angiogenic disease; hyperproliferative disorder;		
KW	KW	diabetic retinopathy; uveitis; rheumatoid arthritis; psoriasis;		
KW	KW	wound healing; endometriosis; vasculogenesis; atherosclerosis;		
KW	KW	nervous system disease; Parkinson's disease; Alzheimer's disease;		
KW	KW	infectious disease; Gene; ss.		
XX	OS	Homo sapiens.		
XX	XX			
PH	Key	Location/Qualifiers		
FT	CDS	112..4731		
FT		/*tag= a		
FT		/product= "Novel secreted protein"		
FT	sig_peptide	112..114		
FT		/*tag= b		
FT	mat_peptide	115..4728		
FT		/*tag= c		
FT		/label= mature_human_secreted_protein		
XX	XX	WO200214341-A1.		
XX	XX	21-FEB-2002.		
XX	XX	13-AUG-2001; 2001WO-US025288.		
XX	XX	14-AUG-2000; 2000US-0225215P.		
XX	PA	(HUMA-) HUMAN GENOME SCI INC.		
XX	PI	Barash SC, Ni J, Ruben SM, Rosen CA, Shi Y;		
XX	XX	WPI; 2002-257592/30.		
DR	DR	P-PSDB; AAU77729.		
XX	XX	Novel human secreted proteins useful for treating immune system e.g.		
PT	PT	anemia, cardiovascular e.g. myocardial infarction, neurological disorders		
PT	PT	e.g. Alzheimer's disease, Parkinson's disease.		
XX	XX	Claim 1; Page 318-320; 337pp; English.		
XX	XX	The invention describes eight novel human secreted polypeptides (I). (I)		
CC	CC	and the polynucleotides encoding them are useful for treating disorders,		
CC	CC	including disorders of immune system e.g. immunologic deficiency		
CC	CC	syndrome, ataxia telangiectasia, HIV infection (human immunodeficiency		
CC	CC	virus), Wiskott-Aldrich disorder, anaemia, thrombocytopenia, heart		
CC	CC	attacks (infarction), strokes, or scarring; autoimmune diseases e.g.		
CC	CC	Addison's disease, encephalomyelitis, multiple sclerosis, autoimmune		
CC	CC	thyroiditis, lupus, erythematosis, insulin dependent diabetes mellitus;		
CC	CC	disorders of inflammation e.g. nephritis, ischaemia-reperfusion injury,		

CC	hypergammaglobulinaemia, Crohn's disease, sarcoidosis, Gaucher's disease;
CC	cardiovascular disease e.g. cardiovascular abnormalities, persistent
CC	truncus arteriosus, arrhythmias, angina pectoris, myocardial infarction,
CC	hypertension, telangiectasia, ischaemia includes cerebral ischaemia,
CC	mucocutaneous lymph node syndrome; disorders and/or conditions associated
CC	with neovascularisation; cancers which involve angiogenesis, benign
CC	tumours, ocular angiogenic diseases, hyperproliferative disorders,
CC	diabetic retinopathy, uveitis, rheumatoid arthritis, psoriasis, delayed
CC	wound healing, endometriosis, vasculogenesis, atherosclerosis; nervous
CC	system diseases, degeneration associated with Parkinson's disease, a
CC	Alzheimer's disease and infectious diseases. This sequence encodes a
CC	novel human secreted protein, described in the invention
XX	
SQ	Sequence 5402 BP; 1823 A; 929 C; 1049 G; 1601 T; 0 U; 0 Other;
Query Match	11.6%; Score 559; DB 6; Length 5402;
Best Local Similarity	60.8%; Pred. No. 1.2e-142;
Matches 989; Conservative	0; Mismatches 520; Indels 18; Gaps 4;
QY	2824 CTTATTCTCTGAAGATGCGCTTCCTCCCATCTCTCATCTCCACTGGTGTAAGAAGG--AGAC 2880
Db	3157 CCTAATGATGAGGACAGTTTGCCGCCCTCTCTGTTGCATCTGGAGAAAGGATCAGTG 3216
QY	2881 TATGCTGTGGAAGAGAACCATCACAGATTCAGTAATGCAGCAGTGTGGAGATGTGGC 2940
Db	3217 CCTGTAGTAGAAGAACATCCATCTCATGACGACATCATTTTGTCTTCTTGAAGGTGAAGC 3276
QY	2941 CCTGACCCACATGTATTTGTTTAAATGCAAAATTTGTGTCATGTTAAATTTGTAAT 3000
Db	3277 TTTCACTCTGTTACATTTTGCTTAAATGCTAACTACTCTGTAATGCAATTCATATT 3336
QY	3001 TATGTGAACAGGAAGTCTGGTGTTCACAAACGAGGAATGCATGCGTGGTCACTCT 3060
Db	3337 TATTCCTCAGACAAATATTGGTACTTTTCAACCAATGGATTGTCATGCTTGGACAGGCA 3396
QY	3061 GAGATAGTCATTCTTCTACAGTGTTTACCGGATGAAAGTGTTTGGCAAAGGATATCTTT 3120
Db	3397 GAATTAATATTTCTATTTGTTATGTTTGCCTAAATGAAGATACTATTCTTAGGACATCTTC 3456
QY	3121 AATCACTTTGTGCAGCTTTATCGGGATGCTCTGCGAGGGAATGTGTGACCACTTGGGA 3180
Db	3457 AGACTATTATTATCACCATATATAAGGATGCTCTTAAAGGAAATATACATAGAAAACCTTGGAC 3516
QY	3181 CATTCCTTTCTCAGTCAAGTTTCCTTTGGGAGTAAGAAACATGTTGATTTCTTATATGTG 3240
Db	3517 AATATACCTTTACTGAGAGTTTCTCAGTAGCAAGATCACGAGAGATTCCCTGTTATT 3576
QY	3241 ACATCTACCTACCAGTCACTCCAAAGACCTAGTACTCCCAACCCCACTTACTTGTGTTGGG 3300
Db	3577 ACACCTACTTTTTCAGAACTTTGATGATCTCTCTCATACCAAGTAATCTCTTTTCTTGTGGA 3636
QY	3301 ATTCTATTCAGAAATGGAACTCCCTTGGGCTAAAGTATTTCCTATCCGCTGATGTTG 3360
Db	3637 ATTCTTATCCAGAGCTTGAGATTCCTCTGGGCAAAAGGTTTTTCCATTCGGTTAAAGTTG 3696
QY	3361 AGACTTGGAGTGAATATCGACTTTTATCCATGCCACCTATTTCAGTGTCAAGATTCGGAAG 3420
Db	3697 AGATTGGGTGCAGAAATAAAGCATATCTCTCTCTCTTAAACAAGCATCAGAGGCCGAAAA 3756
QY	3421 CCATTGTTTGGAGACCGGGCATACATCATGATCTTCTTGCAGACTTTCAGAAATTCAC 3480
Db	3757 CCTCTTTTGGAGAAATAGGACACACTATTATGAACCTTACTTTGTTGACCTTCGAAATATC 3816
QY	3481 CAGTATACCTCTGCCAGTAGTTTCAAGGTTTGGTGGTTGATATGGAAGTTCGGAATAACTPAGC 3540
Db	3817 CAGTATACCTTGCATAATAATAGATCAACTGTTTGATTCATATGGAATAATGGGAAAAAGCTGC 3876
QY	3541 ATCAAAATTCACAGCAACAGATACATAGATGATGAAGCCATGAACAGTCCAAATGAG 3600
Db	3877 ATAAAAATACCAACGGAAAAAGTACAGTGAATTAATGAAGTACAAATTTCTTCCATGAG 3936
QY	3601 CATGTCCTGGCAGGAGGTGCCTGCTTCAATGAAAAGGCAGACTCTCATCTTGTGTGTGA 3660

CC	hypergammaglobulinaemia, Crohn's disease, sarcoidosis, Gaucher's disease;
CC	cardiovascular disease e.g. cardiovascular abnormalities, persistent
CC	truncus arteriosus, arrhythmias, angina pectoris, myocardial infarction,
CC	hypertension, telangiectasia, ischaemia includes cerebral ischaemia,
CC	mucocutaneous lymph node syndrome; disorders and/or conditions associated
CC	with neovascularisation; cancers which involve angiogenesis, benign
CC	tumours, ocular angiogenic diseases, hyperproliferative disorders,
CC	diabetic retinopathy, uveitis, rheumatoid arthritis, psoriasis, delayed
CC	wound healing, endometriosis, vasculogenesis, atherosclerosis; nervous
CC	system diseases, degeneration associated with Parkinson's disease,
CC	Alzheimer's disease and infectious diseases. This sequence encodes a
CC	novel human secreted protein, described in the invention
XX	
XX	Sequence 5402 BP; 1823 A; 929 C; 1049 G; 1601 T; 0 U; 0 Other;
Query Match 11.6%; Score 559; DB 6; Length 5402;	
Best Local Similarity 60.8%; Pred No. 1.2e-142;	
Matches 989; Conservative 0; Mismatches 620; Indels 18; Gaps 4;	
QY	2824 CTTATTCCTGAAGATGGCTTCTCCCAATTCATCTCCACTGGTGTAAAAGG---AGAC 2880
Db	3157 CCTAATGATGAGCAGACATTTGCCCCCACTTGTGGTTCATCTGGAGAAAAGGATCAGTG 3216
QY	2881 TATGCTGTGGAGAGAAACCATCAGATTTTCAGTAATGCAGCAGTTGAGGATGGTGGC 2940
Db	3217 CCTGTAGTAGAAGACATCCATCTCATGAGCAGATCATTTTCTTCTTGAAGTGAAGC 3276
QY	2941 CCTGACCCACATTTGTTTGTAAATGCAATTTTGTCAATGTTTAAATTTTAAATTTTAAAT 3000
Db	3277 TTTTCATCTGTTTACATTTTGTCTTAAATGTAATCTACTCTGTAATGTCAATTTTCAATTT 3336
QY	3001 TATGTGAACAGGAAGTCTGGTGTTCACCAACAGGAATGCATGCAGTGGTGCAGTCT 3060
Db	3337 TATTCCTCAGACAAATATTTGGTACTTTTCAACCAATGGATGTCAGTGGTGGACAGCA 3396
QY	3061 GAGATAGTCATTTCTACAGTGTTTACCGGATGAAAGTGTTCGCCAAGGATATCTTT 3120
Db	3397 GAAATTTATTTCTATTGTTTGTGTTTGCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3456
QY	3121 AATCACTTTTGTCCAGCTTTATCGGATGCTCTGGCAGGGAATGTGGTGAAGCACTTGGGA 3180
Db	3457 AGACTATTTTATCCCATATATATAGGATGCTCTTAAAGGAAAAATACATAGAAAACCTGGAC 3516
QY	3181 CATTCCTTCTCAGTCAAAAGTTTCTTGGCAGTAAGAACATGTTGGATTTCTTATATGTG 3240
Db	3517 AATATTACCTTTACTGAGAGTTTCTCAGTAGCAGGATCAGCGGAGATTCCTGTTTATT 3576
QY	3241 ACATCTACCTACCAGTCACTGCAAGACCTAGTACTCCCAACCCCACTTCTTCTTCTTCTTCT 3300
Db	3577 ACACCTACTTTTTCAGAAACTTGTATGATCTCTCATTTACCAAGTAATCCTTTTCTTCTTCT 3636
QY	3301 ATTCTTTATCCAGAAATGGGAAACTCTTGGGCTAAAGTATTTTCTTCTTCTTCTTCTTCTTCT 3360
Db	3637 ATTCTTTATCCAGAAAGCTTGAGATTCCTTGGGCAAGGTTTTCCTTATGCTTTTAAATGTTG 3696
QY	3361 AGACTTGGAGCTGAATATCGACTTTTATCCATGCCACTATTTCAGTGTCCAGATTTCCGAAAG 3420
Db	3697 AGATTGGTGCAGAAATATTAAGCATATCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3756
QY	3421 CCATTTGTTGGAGAGCGGGGATACCATCATGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3480
Db	3757 CCTCTTTTGGAGAAATAGGACACACTATTATGAACCTTACTTGTGTGACCTTCGAAATTTAC 3816
QY	3481 CAGTATACCTTGGCCAGTAGTTTCAAGTTTGGTGTGATGATGGAAGTTCGGAAGCTAGC 3540
Db	3817 CAGTATACCTTGGCATATATATAGTCACTGTTGATTCATATGAAATGGGAAAAAGCTGC 3876
QY	3541 ATCAAAATTTCCAGCAACAGATACAAATGATGATGAATGAAGCCATGAACAGTCCAAATGAG 3600
Db	3877 ATAAAATATCCACGGAAAAAGTACAGTGATGAATGAAAAGTACTAAATTTCTTCCAAATGAG 3936
QY	3601 CATGTCTGTCAGAGGTCCTGCTTCAATGAAGGCGAGACTCTCATCTTGTGTGTGTA 3660


```
QY 1740 CAGCACTGGTGACCTAATGAGAAATATTTACATAATTTCTGTAGTCAAGTTCCATC 1799
Db 1411 TGGCAAGGTTGACCAATGAAGAACAA---GTTTGCATAGTCTTTAATCCGGTTCATC 1467
QY 1800 AGTGCCTTGGGCAATCTTCCGCCAAGTAGTAGAGCCTGCCATCTATCAGTGTTCCTTT 1859
Db 1468 CATGCATGGGCAACCTCACCAAAAAGGCAAGATTGTGCAATCCCTCAGTGTTCATA 1527
QY 1860 TGGTGGTGAAGACCCAGCAACTTCTTAATCTTAACCTCAATTTCCAAAGCCATTATC 1919
Db 1528 TGGTGGAGCTCGGCCAAGCAGCAACTCATCTCAAACTCAATTTCCACAGCCATTGTC 1587
QY 1920 AGACCAATTTCAAAATCA---CTTCTCTGCAAAACAGTGGAAATAATACTAAAAATAAAA 1976
Db 1588 TGAATGTTACAGTGTGATCTCATTCGCCCAATGCTGGATGAGTCTTAATAACAAAA 1647
QY 1977 TGATATTTCTTGGGAAACAAAATTTAGGGGAAACTCAGCAACCAATGATGCAAGTCCATC 2036
Db 1648 TGACATCTTAAACAAATCAAAATCGGGGGGATAACCTGATTTCAGAATCACTACGTGAGGA 1707
QY 2037 TTTGGGAAACATCTCTAATCTCGATACAAATGGGAAACATTTAGAAAGTTATGAGGCTGA 2096
Db 1708 AGTGCAGGCCCTGTTACTGATACAAATGGTGAAGTCCCTCGAGAAACAGGGGACCTGG 1767
QY 2097 GATCTCCACTAGACCATGCTTGCATTAGCTCCAGATAGCCCAAGATATGATCTCAGAGC 2156
Db 1768 CAGC-----CTGTGSCCTTGCAAGTGTCTCCAGACAGCCCTCACAAATGATCTGCTGC 1818
QY 2157 TGGTCAAGTTGGAAATTTCTGCCAGAAAGCCATTCACCACGCTGGTGAGTGCTCCAGT 2216
Db 1819 TGGACAGTTTGGGTACCCCATCTCTAAGCCATTTACTACTAGGGGATGTGGCTCCAGT 1878
QY 2217 ATGGGTACCGGATTTCTCAGGCTCCAAATTTGCATGAAATGTGAAGCCAGGTTTACATTCAC 2276
Db 1879 CTGGGTGCCAGATTCCCAAGCACCACCACTGCATGAAGTGGAGGCCAGATTTACATTTAC 1938
QY 2277 CAAAAGGAGGCATCACTGCAGAGCATGTGGGAAGGTTT 2314
Db 1939 CAAAAGGAGGCATCACTGCCGAGCTGTGGAAAGGTAT 1976
```

```
RESULT 15
AAC30390
ID AAC30390 standard; cDNA; 332 BP.
AC AAC30390;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 34465.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-00200610.
XX
PR 26-FEB-1999; 99US-0122487P.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
```

```
XX Claim 1; SEQ ID NO 34465; 71pp + Sequence Listing; English.
PS
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. NO ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors
XX
SQ Sequence 332 BP; 105 A; 67 C; 70 G; 90 T; 0 U; 0 Other;
Query Match 6.6%; Score 318; DB 3; Length 332;
Best Local Similarity 99.7%; Pred. No. 7.5e-77;
Matches 329; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 171 CAAGGTGTTAGATGAATTTGAACAAACGAGATGAACAGTTTCTTCTACTTTATTGGA 230
Db 1 CAAGGTGTTAGATGAATTTGAACAAACGAGATGAACAGTTTCTTCTACTTTATTGGA 60
QY 231 TACAAAGTGGAAATAGATCTAGATCCGCCCTTCTCACCGGCTGTGCAATTTAACTTCTT 290
Db 61 TACAAAGTGGAAATAGATCTAGATCCGCCCTTCTCACCGGCTGTGCAATTTAACTTCTT 120
QY 291 GGCC-AGTGTGAATGAATCTGCAGTTTCTAATGAGTCAACACCACTGAAAGTCTTCT 349
Db 121 GGCCAAAGTGTGAATGAATCTGCAGTTTCTAATGAGTCAACACCACTGAAAGTCTTCT 180
QY 350 CCTGCTCATTTTCAGTCCCTTGACACAGAGGAAAGAGGATCACTGTGTAATGGACAGG 409
Db 181 CCTGCTCATTTTCAGTCCCTTGACACAGAGGAAAGAGGATCACTGTGTAATGGACAGG 240
QY 410 ACTGTAATCTAAATCCAGAGATTGCCCAATGTGAATGATGAAAATGCTGTTCGAGAAG 469
Db 241 ACTGTAATCTAAATCCAGAGATTGCCCAATGTGAATGATGAAAATGCTGTTCGAGAAG 300
QY 470 ACCAGTTAATTAAGAGAACTATAGTTGGG 499
Db 301 ACCAGTTAATTAAGAGAACTATAGTTGGG 330
```

Search completed: May 17, 2004, 18:15:06
Job time : 1124.1 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 16:26:01 ; Search time 7415.64 Seconds
(without alignments)
19486.266 Million cell updates/sec

Title: US-09-744-167-1
Perfect score: 4839
Sequence: 1 gcatactgaatcagcaggac.....atcaaaaaaaaaaaaaaaaaaa 4839

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_estum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_est1: *
10: gb_est2: *
11: gb_htc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: em_gss_hum: *
18: em_gss_inv: *
19: em_gss_pin: *
20: em_gss_vrt: *
21: em_gss_fun: *
22: em_gss_mam: *
23: em_gss_mus: *
24: em_gss_pro: *
25: em_gss_rtd: *
26: em_gss_phg: *
27: em_gss_vrl: *
28: gb_gss1: *
29: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	919.8	19.0	1010	12	BM463429
2	831.8	17.2	1047	13	BU745832
3	731.4	15.1	750	12	BI333812
4	714.2	14.8	918	13	BQ952036

C	5	699.2	14.4	706	12	BM674358
C	6	699.2	14.4	941	13	BQ421818
C	7	656.2	13.6	673	10	BF433490
C	8	648.6	13.4	751	12	BI334801
C	9	646	13.3	669	12	BM716723
C	10	636.2	13.1	921	14	CF408646
C	11	634	13.1	694	14	CB047954
C	12	631.4	13.0	644	13	BX096354
C	13	628.4	13.0	654	14	CB049545
C	14	611.2	12.6	777	14	CA322478
C	15	609.4	12.6	806	13	BU745833
C	16	599	12.4	906	10	BF671658
C	17	587	12.1	754	12	BI149584
C	18	576.4	11.9	657	12	BG536226
C	19	570.8	11.8	610	14	H97939
C	20	570	11.8	612	10	BE364762
C	21	566.4	11.7	568	9	AI632587
C	22	560.6	11.6	734	12	BM963298
C	23	559.4	11.6	569	13	BQ340247
C	24	556	11.5	565	14	CB047955
C	25	554.4	11.5	598	13	BQ340252
C	26	550.8	11.4	592	10	AW614993
C	27	543.6	11.2	606	10	BE588313
C	28	542.6	11.2	604	9	AI207521
C	29	540.6	11.2	955	12	BG676689
C	30	540.2	11.2	884	14	CF408647
C	31	531	11.0	539	12	BI035332
C	32	530.4	11.0	541	10	AW301671
C	33	522.8	10.8	558	10	BF189106
C	34	519.4	10.7	595	14	CB048575
C	35	487.4	10.1	534	10	BE078439
C	36	476.8	9.9	522	10	AW580326
C	37	471.8	9.7	498	10	BF090054
C	38	466.4	9.6	510	10	BF953986
C	39	455	9.4	467	9	AI797623
C	40	450.8	9.3	642	13	BQ339364
C	41	449.4	9.3	544	13	BQ339320
C	42	445.4	9.2	453	9	AA156466
C	43	444	9.2	460	10	AW377196
C	44	443.8	9.2	791	13	BE872429
C	45	443	9.2	443	10	BE048597

ALIGNMENTS

RESULT 1
BM463429
LOCUS
DEFINITION
AGENCOURT_6433518 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5585817
5', mRNA sequence.
ACCESSION
BM463429
VERSION
EST.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1010)
NIH-MGC http://mgc.nci.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: ATCC
cdna Library Preparation: Life Technologies, Inc.
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LHAM12352 row: 1 column: 10
High quality sequence stop: 666.

BM463429 1010 bp mRNA linear EST 05-FEB-2002
AGENCOURT_6433518 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5585817
5', mRNA sequence.

ACCESSION
BM463429

VERSION
EST.

KEYWORDS
Homo sapiens (human)

SOURCE
Homo sapiens

ORGANISM
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 1010)

NIH-MGC http://mgc.nci.nih.gov/

TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
Unpublished (1999)

COMMENT
Contact: Robert Strausberg, Ph.D.

Email: cgapsb@mail.nih.gov

Tissue Procurement: ATCC

cdna Library Preparation: Life Technologies, Inc.

cdna Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov

Plate: LHAM12352 row: 1 column: 10

High quality sequence stop: 666.

Db 973 AGATCTCAGGCTGCACTCCATGAAATGTGAA-CCAGGGTCACATTCACCNAAAGGAG 915
Qy 2286 GCATCACTGCAGACATGTGGGAAGTTTCTGTGCTTCTGCTGTAGCCTGAAATGTAA 2345
Db 914 GCATCACTGCAGACATGTGGGAAGTTTCTGTGCTTCTGCTGTAGCCTGAAATGTAA 855
Qy 2346 ACTGTTATACATGACAGAAAGAGTGTAGTGTGTGTATCTGCCATTCAGTGTCTAAT 2405
Db 854 ACTGTTATACATGATAGAAAGGAGCTAGATGTGTGTATCTGCCATTCAGTGTCTAAT 795
Qy 2406 GAATGCTCAAGCTCGGAGAACATGATGAGTGCTCAAGCCAGAGCCCTAACCTTAACAA 2465
Db 794 GAATGCTCAAGCTCGGAGAACATGATGAGTGCTCAAGCCAGAGCCCTAACCTTAACAA 735
Qy 2466 TCCTGCTGAATATCTTCTACTATCCCTCCCTTCAGCAAGCTCAGCCCTCAGAGCTCT 2525
Db 734 TCCTGCTGAATATCTTCTACTATCCCTCCCTTCAGCAAGCTCAGCCCTCAGAGCCCT 675
Qy 2526 GAGCTCTCCACTCCCTGATGTGATGTACCTGTGGGAGTTTAAAGCACCTCGAGCAGA 2585
Db 674 GAGCTCTCCACTCCCTGATGTGATGTACCTGTGGGAGTTTAAAGCACCTCGAGCAGA 615
Qy 2586 AGTGGCTCAGCCAGAGAGAGAGCGAGTTTGGTTTCTGATGGGATCTTGCCCAATGG 2645
Db 614 AGTGGCTCAGCCAGAGAGAGAGCGAGTTTGGTTTCTGATGGGATCTTGCCCAATGG 555
Qy 2646 AGAAGTTGCTGATGAGCGCAATTAACATGAATGGAATCTCTCTCAGGACCTCGGC 2705
Db 554 AGAAGTTGCTGATGAGCGCAATTAACATGAATGGAATCTCTCTCAGGACCTCGGC 495
Qy 2706 TGTGTCAACAGCCAGTCAAGCCAGTAACTPACAGTCTCTTACAGCAGAGAGCGGATAT 2765
Db 494 TGTGTCAACAGCCAGTAAAGCCAACTAACAGTCTCTTACAGCAGAGAGCGGATAT 435
Qy 2766 TTGTCTATCTCTGGGATATACTCAGGTTGGAGTCTCTGTTGGAAGTGCAATGAATCT 2825
Db 434 TTCTCTGTCTCTGGGATATACTCAGGTTGGAGTCTCTGTTGGAAGTGCAATGAATCT 375
Qy 2826 TATTCCTGAAGTGGCTCTCTCCATCTCTCATCTCCACTGCTGTAAGAGAGACTATGC 2885
Db 374 AATTCCTGAAGTGGCTCTCTCCATCTCTCATCTCCACTGCTGTAAGAGAGACTATGC 315
Qy 2886 TGTGGAAGAGAAACCATCACAGATTCAGTAATGAGAGTGGAGATGGTGGCCCTGA 2945
Db 314 TGTGGAAGAGAGCCATCACAGATTCAGTAATGAGAGTGGAGATGGTGGCCCTGA 255
Qy 2946 CCCACTGTTATTTTAAATGCAATTTGTTGTCATGTTTAAATTTAAATTTAATTTATG 3005
Db 254 CCCACTGTTATTTTAAATGCAATTTGTTGTCATGTTTAAATTTAAATTTAATTTATG 195
Qy 3006 GAACAGGAAGTCTGCTGTTTCAACCAAGGGAATGATGAGTGGTCTAGTCTGAGAT 3065
Db 194 GAACAGGAAGTCTGCTGTTTCAACCAAGGGAATGATGAGTGGTCTAGTCTGAGAT 135
Qy 3066 AGTCATCTTCTPACAGTGTTCACCGATGAAAGTGTGTTGCAAGAGATATCTTTAAT 3123
Db 134 AGTCATCTTCTGCAAGTGTCTACCGATGAAAGTGTGTTGCAAGAGATATCTTTAAT 77

RESULT 3
BI333812 750 bp mRNA linear EST 30-JUL-2001
LOCUS 602999315F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5141571 5',
DEFINITION mRNA sequence.
ACCESSION BI333812
VERSION BI333812.1 GI:15018469
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 750)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11348 row: n column: 04
High quality sequence stop: 745.
FEATURES
Location/Qualifiers
source 1..750
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5141571"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/clone_lib="NIH_MGC_12"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

ORIGIN

Query Match 15.1%; Score 731.4; DB 12; Length 750;
Best Local Similarity 99.7%; Pred. No. 3.9e-134; Indels 1; Gaps 1;
Matches 743; Conservative 0; Mismatches 1;
Qy 3238 GTGACATCTACCTACCTGCAAGACCTAGTACTCCCAACCCACCTTACTTGT 3297
Db 1 GTGACATCTACCTACCTGCAAGACCTAGTACTCCCAACCCACCTTACTTGT 60
Qy 3298 GGAATCTTATCCAGAAATGGGAACTCTTGGCTAAAGTATTTCTATCGTCTGATG 3357
Db 61 GGAATCTTATCCAGAAATGGGAACTCTTGGCTAAAGTATTTCTATCGTCTGATG 120
Qy 3358 TTGAGACTTGGAGCTCAATATGACATTCATGCCCCTACTATTCAGTGTGAGATTCGG 3417
Db 121 TTGAGACTTGGAGCTCAATATGACATTCATGCCCCTACTATTCAGTGTGAGATTCGG 180
Qy 3418 AAGCCATTTTGGAGAGACGGGGCATACATCATGAAATCTTCTTCCAGACTTCAGAAAT 3477
Db 181 AAGCCATTTTGGAGAGACGGGGCATACATCATGAAATCTTCTTCCAGACTTCAGAAAT 240
Qy 3478 TACAGTATACCTTGGCAGTGTTCAGGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGT 3537
Db 241 TACAGTATACCTTGGCAGTGTTCAGGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGT 300
Qy 3538 AGCATCAAAATTTCCAGCAACAGATACAAATGAGATGATGAAAGCCATGAACAGTCCAAT 3597
Db 301 AGCATCAAAATTTCCAGCAACAGATACAAATGAGATGATGAAAGCCATGAACAGTCCAAT 360
Qy 3598 GAGCATGTCTTGGCAGAGTGGCTCTCTCAATGAAAGGAGACTCTCATCTTGTGTGT 3657
Db 361 GAGCATGTCTTGGCAGAGTGGCTCTCTCAATGAAAGGAGACTCTCATCTTGTGTGT 420
Qy 3658 GTACAGATGATGATGAAACTATCAGCCAGGCTATCAGTATTCAGTATTCAGTATTCAGCC 3717
Db 421 GTACAGATGATGATGAAACTATCAGCCAGGCTATCAGTATTCAGTATTCAGTATTCAGCC 480
Qy 3718 AAAGTGAATGGTCCAGTGTTCCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3777
Db 481 AAAGTGAATGGTCCAGTGTTCCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
Qy 3778 CTTGCCAAGTCCAGTATTTGGAGATGGTGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3837
Db 541 CTTGCCAAGTCCAGTATTTGGAGATGGTGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 600

```

QY 3838 GATTCCTTGA-GGCAGGCACTCGAGAGATGAAGCACTTCAACATCACTCGGTGGGAAGGC 3896
DB 601 GATTCCTTGGGCGCAGGCACTCGAGAGATGAAGCACTTCAACATCACTCGGTGGGAAGGC 660
QY 3897 GGACGGGAGGAAACCCAGGACACATCCACATCCAGTGGGTGGATGATGACAGACGT 3956
DB 661 GGACGGGAGGAAACCCAGGACACATCCACATCCAGTGGGTGGATGATGACAGACGT 720
QY 3957 TAGCAAGGGTGTGCTGAAGTCTTATA 3981
DB 721 TAGCAAGGGTGTGCTGAAGTCTTATA 745

RESULT 4
BQ952036 918 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT 8953735 NCI CGAP Co24 Mus musculus cDNA clone
DEFINITION IMAGE:6477813 5', mRNA sequence.
ACCESSION BQ952036
VERSION BQ952036.1 GI:22367514
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CNA Library Preparation: Life Technologies, Inc.
CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14020 row: f column: 22
High quality sequence stop: 716.
Location/Qualifiers
1. .918
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6477813"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Co24"
/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

FEATURES
source
Query Match 14.8%; Score 714.2; DB 13; Length 918;
Best Local Similarity 88.2%; Pred. No. 9.5e-131;
Matches 812; Conservative 0; Mismatches 103; Indels 6; Gaps 3;

QY 3607 CTGGCAGGAGTGGCTGTTCAATGAAGGACAGCTCTCATCTTGTTGTGTACAGAAT 3666
DB 2 CTGGCAGGAGTGGCTGTTCAATGAAGGACAGCTCTCATCTTGTTGTGTACAGAAT 61
QY 3667 GATGATGAAACTATCAGACCCAGGCTATCAGTATTCACATCAGCCAGCAAGAGTGACT 3726
DB 62 GATGATGAAACTATCAGACCCAGGCTATCAGTATTCACATCAGCCAGCAAGAGTGACT 121
QY 3727 GGTGCCAGTTCTTGTGTTTCACTGGGGCTCTGAATCCCTCTTCTGTATACCTTGCCAAG 3786
DB 122 GGCGCCAGTTCTTGTGTTTCACTGGGGCTCTGAATCCCTCTTCTGTATACCTTGCCAAG 181
QY 3787 TCCAGTATTGTGGAAGTGGTGTATGGTCCAGATTACTCAGAGAACATGGATTCTCTTG 3846

```

```

DB 182 TCCAGTATTGTGAAGATGGAGTCATGGTCCAGATCACTGCGGAGAACATGGACTCTTTG 241
QY 3847 AGCAGGCACCTGCGAGAGATGAAGCACTTCAACATCACTGTGGGAAGCGCAGCGGAG 3906
DB 242 AGCAGGCACCTGCGGAGATGAAGCACTTCAACATCACTGTGGGAAGCGGATGCAGAG 301
QY 3907 GAACCCAGAGACATCCACATCCAGTGGGTGGATGATGAACAAGAACTTTAGCAAGGT 3966
DB 302 GACCCCGAGGAAACAGATCCACATCCAGTGGGTGGAGCATGACAAGACAGTTAAACAAGT 361
QY 3967 GTCGTAAGTCCCTATAGATGGGAGTCCATGGAGACTATTAACAATGTGAAGATATTCAT 4026
DB 362 GTGGTGAATCCCTATAGATGGGAGTCTTGGAAATCTTAACAACAGTGAATAATTCAT 421
QY 4027 GGATCAGAAATATAAGCAAAATGGAAGTAATCAGATGGACAGAGGTGTTTTCTAGAA 4086
DB 422 GGATCAGAGTATAAGCAAAACGGAAGTCAACAGATGGACAGAGGTGTTTTCTGGAA 481
QY 4087 AAGCATGACGACACAAATTCCTCAGTGTATCTGCAGATCAGATCAGATGATGATGATGAT 4146
DB 482 AAGCATGACGACACAAATTCCTCAGTGTATCTGCAGTGTATCAGATCAGATGATGATGAT 541
QY 4147 GTTGCCAAAGCTTTTGGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4206
DB 542 GTTGCCAAAGCTTTTGGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 601
QY 4207 ATGACCAAACTGGGACTAGTGTGATCAGTGTATCAGTGTATCAGTGTATCAGTGTATCAG 4266
DB 602 ATGACCAAACTGGGACTAGTGTGATCAGTGTATCAGTGTATCAGTGTATCAGTGTATCAG 661
QY 4267 AGCAATGGCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4326
DB 662 AGCAATGGCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 721
QY 4327 GT-GATCCATGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4385
DB 722 GTCTCTTTCATGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781
QY 4386 TTATATCTTGGAAACATCGTATTAACAGAGAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCT 4445
DB 782 TTATATCTTGGAAA-----CATCGCATAGTAGAGGACTTCCATGTTCTGTTTCAGACTTG 837
QY 4446 TTGCAACAGCAGTCAATCCCAATCATTTGCACTTTAAAACT-GGAAGATTAAAGCTTTTG 4504
DB 838 GTGCGACGACGATCATATCCAAATCATTTGCACTTTAAAACTGCGGAGGAATTAGCTTTTG 897
QY 4505 TTAAACACTATTAAATGGGTGG 4525
DB 898 TTAACGGTCCCATTAAGGGGG 918

RESULT 5
BM674358/ 706 bp mRNA linear EST 27-FEB-2002
LOCUS UI-E-EUJ-ahk-b-16-0-UI.s1 UI-E-EUJ Homo sapiens cDNA clone
DEFINITION UI-E-EUJ-ahk-b-16-0-UI 3', mRNA sequence.
ACCESSION BM674358
VERSION BM674358.1 GI:18984256
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 706)
Nucleotide, M.F., Lennon, G. and Soares, M.B.
Normalisation and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL 97044477
MEDLINE 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa

```

375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 Forward
POLYA-yes

Location/Qualifiers
1. 706
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EJO-ahk-b-16-0-UI"
/tissue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"
/dev stage="fetal and adult"
/lab host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EJO"
/note="Organ: eye; Vector: pTV73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-EJO is a subcloned cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pTV73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment, AATGCCGAT; optic nerve, CGATTAAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene discovery in the Visual System, supported by National Eye Institute (NEI). TAG_TISSUE=Foveal and Macular Retina TAG_LIB=UI-E-EJO TAG_SEQ=GTCC"

ORIGIN
Query Match 14.4%; Score 699.2; DB 12; Length 706;
Best Local Similarity 99.4%; Pred. No. 9.1e-128;
Matches 701; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 950 AGGGACAGATGGATGCTCTGTTTAAAGCAAGAGAACTATATACGATGAGGACC 1009
DB 706 AGGGACAGATGGATGCTCTGTTTAAAGCAAGAGAACTATATACGATGAGGACC 647
QY 1010 TCATGGCAAAATCAGCTCTCTAGGACAGATCTAGGAGTCCAAATTCCTTTCCAC 1069
DB 646 TCATGGCAAAATCAGCTCTCTAGGACAGATCTAGGAGTCCAAATTCCTTTCCAC 587
QY 1070 TCAGTGAGGGATTTTGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1129
DB 586 TCAGTGAGGGATTTTGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 527
QY 1130 TCCGCTCTGTTTACCTTTGCTTCTCAAAACAGAGATCCCTAATGGGCTCGAAGGA 1189
DB 526 TCCGCTCTGTTTACCTTTGCTTCTCAAAACAGAGATCCCTAATGGGCTCGAAGGA 467
QY 1190 ATGACTGTGAACGGTTCAGATTCCTTGTCCCTAATGCAAGTTCAGGCTCATGAATG 1249
DB 466 ATGACTGTGAACGGTTCAGATTCCTTGTCCCTAATGCAAGTTCAGGCTCATGAATG 407
QY 1250 AAGGTTATCAACATCAAGAAACTCTTGGCACTACAGAAATTCCTTAATATGACAGGCA 1309
DB 406 AAGGTTATCAACATCAAGAAACTCTTGGCACTACAGAAATTCCTTAATATGACAGGCA 347

QY 1310 TCTCTGAATCTCAGGACATGACTAATTTGGAAGTTGACTAACTAAATGAGATGATGATA 1369
DB 346 TCTCTGAATCTCAGGACATGACTAATTTGGAAGTTGACTAACTAAATGAGATGATGATA 287
QY 1370 GCCAAGTAAACCAAGAAAGGAAAAAGTTTCTACAGATTAGTCAGCCTGAGGACACTAATG 1429
DB 286 GCCAAGTAAACCAAGAAAGGAAAAAGTTTCTACAGATTAGTCAGCCTGAGGACACTAATG 227
QY 1430 GTGATAGTGGAGGACAGCTGTGTGGATTGGCAGATGCGAGTCTAGATTAAAGGAATTT 1489
DB 226 GTGATAGTGGAGGACAGCTGTGTGGATTGGCAGATGCGAGTCTAGATTAAAGGAATTT 167
QY 1490 GCATTAGTGAAGAGTGAAGATGCTGATTTCTCCACTGTTTATAGACACACAGCAGCAAAAT 1549
DB 166 GCATTAGTGAAGAGTGAAGATGCTGATTTCTCCACTGTTTATAGACACACAGCAGCAAAAT 107
QY 1550 ATCTATCTAATCGTTGTGATTCCTATGGAATGCAAGACCAGGCTGTTCTTTTGTCCAA 1609
DB 106 ATCTATCTAATCGTTGTGATTCCTATGGAATGCAAGACCAGGCTGTTCTTTTGTCCAA 47
QY 1610 AGACTTTCCTCCCAAGAGAGATTTCAGTAACAGAGAAAAAGAAA 1654
DB 46 AGACTTTCCTCCCAAGAGAGATTTCAGTAACAGAGAAAAAGAAA 2

RESULT 6
LOCUS BQ421818 941 bp mRNA linear EST 23-MAY-2002
DEFINITION AGENCOURT_7761822 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6018276
5', mRNA sequence.
ACCESSION BQ421818
VERSION BQ421818.1 GI:21117133
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 941)
AUTHORS NIH-MGC http://img.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13218 row: o column: 13
High-quality sequence stop: 609.
Location/Qualifiers
1. 941
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6018276"
/tissue_type="epithelioid carcinoma"
/lab host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 70"
/note="Organ: pncrfas; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.1 kb. Library constructed by Life Technologies."

ORIGIN
Query Match 14.4%; Score 699.2; DB 13; Length 941;
Best Local Similarity 98.0%; Pred. No. 8.8e-128;
Matches 729; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

QY 3433 GAGCGGGGATACCATCATCAATCTCTTTCGAGAGTTTCAGAAATTCACGATACCTTG 3492

IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 471.

FEATURES
source

[illegible][illegible]

```

Qy 4761 TGGGTACCAACCATCACCAAGGGTGGGATGGAGGGGAGGGGAGGGAATAATAAAGCA 4820
Db 73 TGGGTACCAACCATCACCAAGGGTGGGATGGAGGGGAGGGGAGGGAATAATAAAGCA 14
Qy 4821 TCAAAAAAAAAA 4833
Db 13 TCAAAAAAAAAA 1

RESULT 8
LOCUS BI334801 751 bp mRNA linear EST 30-JUL-2001
DEFINITION 60298915F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5141187 5',
mRNA sequence.
ACCESSION BI334801
VERSION BI334801.1 GI:15019458
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
1 (bases 1 to 751)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11347 row: n column: 04
High quality sequence stop: 742.
FEATURES
    source
        location/Qualifiers
            1..751
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:5141187"
                /tissue_type="cervical carcinoma cell line"
                /lab_host="DH10B"
                /clone_lib="NIH_MGC_12"
                /note="Organ: Cervix; Vector: pCMV-SPORT6; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.4 kb. Library prepared by Life
                Technologies."
ORIGIN
Query Match 13.4%; Score 648.6; DB 12; Length 751;
Best Local Similarity 94.3%; Pred. No. 9.1e-118;
Matches 708; Conservative 0; Mismatches 34; Indels 9; Gaps 3;

Qy 3238 GTGACATCTACTACGACTCACTCAAGACCTAGTACTCCCAACCCACCTTACTTGT 3297
Db 1 GTGACATCTACTACGACTCACTCAAGACCTAGTACTCCCAACCCACCTTACTTGT 60
Qy 3298 GGGATCTTATCCAGAAATGGGAACCTCTTGGGCTAAAGTATTCTTCCTCGTCTGATG 3357
Db 61 GGGATCTTATCCAGAAATGGGAACCTCTTGGGCTAAAGTATTCTTCCTCGTCTGATG 120
Qy 3358 TTGAGACTTGGAGCTGAATATCGATCTTATCCATGCCACCTATTCAAGTGTGAGATTTCGG 3417
Db 121 TTGAGACTTGGAGCTGAATATCGATCTTATCCATGCCACCTATTCAAGTGTGAGATTTCGG 180
Qy 3418 AAGCCATTGTTGGAGACGGGGGCATACCATCATGAATCTTCTTCGAGACTTCAGAAAT 3477
Db 181 AAGCCATTGTTGGAGACGGGGGCATACCATCATGAATCTTCTTCGAGACTTCAGAAAT 240
Qy 3478 TACCACTATACCTCCAGTAGTTCAAGGTTGGTGGTGTGATGGAGTTCCGAAACT 3537

```

```

Db 241 TACCAGTATACCTTCCAGTAGTTCAAGTTTGGTGGTGTGATAGAAAGTTCGGAAGACT 300
Qy 3538 AGCATCAAAATTCACGCAACAGATACAATGAGATGATGAAGCCATGAACAGTCCAAAT 3597
Db 301 AGCATCAAAATTCACGCAACAGATACAATGAGATGATGAAGCCATGAACAGTCCAAAT 360
Qy 3598 GAGCATGCTCTGGCAGGAGGTCCTGCTTCAATGAAAAGGAGGAGCTCTCATCTTGTGT 3657
Db 361 GAGCATGCTCTGGCAGGAGGTCCTGCTTCAATGAAAAGGAGGAGCTCTCATCTTGTGT 420
Qy 3658 GTACAGATGATGATGGAAGTATATCAGACCCAGGCTATCAGTATTCAAAATCAGCCAGA 3717
Db 421 GTACAGATGATGATGGAAGTATATCAGACCCAGGCTATCAGTATTCAAAATCAGCCAGA 480
Qy 3718 AAAGTACTGCTGCGCAGTTCTTT--GTGTTCAAGTGGGCTCTG-AAATCCTCTCTCGA 3774
Db 481 AAAGTACTGCTGCGCAGTTCTTTCTCGTGTTCAGTGGGCTCTGCAAAATCCTCTCTCGA 540
Qy 3775 TACCTTGGCAAGTCCAGTATTGTG-----GAAGATGGTGTATTGTCGCAGATTACTGCA 3828
Db 541 TACCTTGGCAAGTCCAGTATTGTGTTGGAACGATGCTGTCATCGTCCACATTACTGCA 600
Qy 3829 GAGAACTGATGATCTCTTGGAGGAGGACTGCGAGAGATGAAGACTTACCATCACCTGT 3888
Db 601 GAGAACTGATGATCTCTTGGAGGAGGACTGCGAGAGATGAAGACTTACCATCACCTGT 660
Qy 3889 GGGAGGCGGAGCGGAGGAGAACCCAGGAGCACATCCACATCCAGTGGGTGATGATGAC 3948
Db 661 GGGAGGCGGAGCGGAGGAGAACCCAGGAGCACATCCACATCCAGTGGGTGATGATGAC 720
Qy 3949 AAGAACTTGAAGAGGGTGTCTGAAGTCTTA 3979
Db 721 AAGAACTTGAAGAGGGTGTCTGAAGTCTTA 751

RESULT 9
LOCUS BI716723 669 bp mRNA linear EST 28-FEB-2002
DEFINITION UI-B-EJO-abk-b-16-0-UI.12 UI-E-EJO Homo sapiens cDNA clone
UI-B-EJO-abk-b-16-0-UI 5', mRNA sequence.
ACCESSION BI716723
VERSION BI716723.1 GI:19029981
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 669)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL 97044477
MEDLINE 8889548
PUBMED
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse
    location/Qualifiers
        1..669
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
FEATURES
    source

```

/clone="UI-E-EJ0-ahk-b-16-0-UI"
 /tissue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"
 /dev_stage="fetal and adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-EJ0"
 /note="Organ: eye; Vector: pT73-pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-EJ0 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dr)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment, AATGCGCAT; optic nerve, CCATTAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Query Match 13.3%; Score 646; DB 12; Length 669;
 Best Local Similarity 99.6%; Pred No. 3e-117;
 Matches 668; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 937 GTAATCTCATCCAGGGAACAGATGGATGCTCTGCTGTTTAAAAAGCAGAGAACTATATA 996
 Db 1 GTAATCTCATCCAGGGAACAGATGGATGCTCTGCTGTTTAAAAAGCAGAGAACTATATA 60

QY 997 CCAGATGAGGACCTCACTGGGAAATCAGCTCTCTAGGACAGATCTAGGAGTCCCAAT 1056
 Db 61 CCAGATGAGGACCTCACTGGGAAATCAGCTCTCTAGGACAGATCTAGGAGTCCCAAT 120

QY 1057 TCCTTTTCCCATCAGTACGAGGGATTTTGATGAAAAAGAGCCAGAGAGAGAGCACC 1116
 Db 121 TCCTTTTCCCATCAGTACGAGGGATTTTGATGAAAAAGAGCCAGAGAGAGAGCACC 180

QY 1117 ACTGAAGATCCCTCGGTCTGTTTACCTTGTCTTCTCAACAGACATGCTTAATGGG 1176
 Db 181 ACTGAAGATCCCTCGGTCTGTTTACCTTGTCTTCTCAACAGACATGCTTAATGGG 240

QY 1177 TCTGAAGAAATATGACTGTGAACGGTGTTCAGATTGCTTGTGCTTAATGAAGTTAGG 1236
 Db 241 TCTGAAGAAATATGACTGTGAACGGTGTTCAGATTGCTTGTGCTTAATGAAGTTAGG 300

QY 1237 GCTGATGAATGAAGTTTGAACATGAAGAACTCTTGCGCATCAGAAATTCCTTAAT 1296
 Db 301 GCTGATGAATGAAGTTTGAACATGAAGAACTCTTGCGCATCAGAAATTCCTTAAT 360

QY 1297 ATGACAGAGCATTTCTCTGAATCTCAGACATGACTAATTCGAAGTTGACTTAACTAAAT 1356
 Db 361 ATGACAGAGCATTTCTCTGAATCTCAGACATGACTAATTCGAAGTTGACTTAACTAAAT 420

QY 1357 GAGATGAATGATAGCCAAAGTAAACGAAAGAAAGAAAGTTTCTACAGATTAGTCAGCCT 1416
 Db 421 GAGATGAATGATAGCCAAAGTAAACGAAAGAAAGAAAGTTTCTACAGATTAGTCAGCCT 480

QY 1417 GAGGACATTAATGTTAGTGGAGGACAGTCTGTGGATGGCAGATGAGGTCTAGAT 1476
 Db 481 GAGGACATTAATGTTAGTGGAGGACAGTCTGTGGATGGCAGATGAGGTCTAGAT 540

QY 1477 TTAAGAGCACTTCATTAAGTCAAGATGAGAAATGTGATTTCTCCACTGTTATAGACACA 1536
 Db 541 TTAAGAGCACTTCATTAAGTCAAGATGAGAAATGTGATTTCTCCACTGTTATAGACACA 600

QY 1537 CCAGCAGCAAAATTAATCTAATGTTGTGATTTCTTATGGAATGCAAGACCCAGGTGT 1596
 Db 601 CCAGCAGC-AAATTAATCTAATGTTGTGATTTCTTATGGAATGCAAGACCCAGGTGT 658

QY 1597 TCCTTTGTTC 1607
 Db 659 TCCTTTGTTC 669

RESULT 10
CF408646/cLOCUS
DEFINITION

CF408646 921 bp mRNA linear EST 02-SEP-2003
 Canis familiaris cDNA clone CH3#055_H02 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

FEATURES

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

ORIGIN

Query Match 13.1%; Score 636.2; DB 14; Length 921;
 Best Local Similarity 86.9%; Pred. No. 2.5e-115;
 Matches 724; Conservative 0; Mismatches 103; Indels 6; Gaps 2;

QY 797 AGGGGAGATCTGTTAAACCATCTGTGTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 856
 Db 892 AGTGGGATCTGTTAAACATATGTGTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 833

QY 857 CCCCTTCACAAATTAAGATGACGAAGTATAGGTAGAGCCCTCCATGCTGGATTA 916
 Db 832 CCCCTTCACAAATTAAGATGATGATACACAGGTAGAGCCCTCCATGCTGGATTA 773

QY 917 CAAGTTTAAACGGTTCATTACAGTAATCTCTATCCAGGGAACAGATGATGATGCTGTTA 976
 Db 772 CAAGTTTAAACATTTGATTTCAGTAATCTCTATCCAGGGAACAGATGATGATGCTGTTA 713

QY 977 AAAGCAGAGAACTATATACAGATGAGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1036
 Db 712 AAAGCAGAGAACTATATATTCAGATGAAGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 653


```

VERSION BX096354.1 GI:27842763
SOURCE EST.
ORGANISM Homo sapiens (human)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M.,
JOURNAL Radelof, U., Schneider, D. and Korn, B.
COMMENT Human Unigeneset - RZPD3
Unpublished (2003)
Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGP998L141206.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
http://www.rzpd.de/CloneCards/cgi-
bin/showlib.pl.cgi/response?libNo=972
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTACACAGAAACAGCATGAC.
FEATURES
source
1..644
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGP998L141206 ; IMAGE:505501"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Soares pregnant uterus NbHPU"
/site="Organ: uterus; Vector: pT7T3-Pac; Site:1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer 15',
AACTGGAGAATTCGCGCCGCGCTTTTTTTTTTTT 3',
(double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
ORIGIN
Query Match 13.0%; Score 631.4; DB 13; Length 644;
Best Local Similarity 99.1%; Pred. No. 2.4e-114;
Matches 632; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 940 ATCTCATCCAGGACAGATGGATGCTCTCTGTTAAAGACGAGACATATATACCA 999
DB 7 ATCTCATCCAGGACAGATGGATGCTCTCTGTTAAAGACGAGACATATATACCA 66
QY 1000 GATGAGGACCTCAGCTGGCAAAATCAGCTCTCTAGGACAGATCTAGGAGTCCAAATTC 1059
DB 67 GATGAGGACCTCAGCTGGCAAAATCAGCTCTCTAGGACAGATCTAGGAGTCCAAATTC 126
QY 1060 TTTTCCACATGATGAGGGGATTTGATGAAAAAGACGAGAGGAGGAGCACCCT 1119
DB 127 TTTTCCACATGATGAGGGGATTTGATGAAAAAGACGAGAGGAGGAGCACCCT 186
QY 1120 GAAGAATCCCTCCGCTCTGTTTACCTTTGCTTCTCAACAGACATGCTTAATGGTCT 1179
DB 187 GAAGATCCCTCCGCTCTGTTTACCTTTGCTTCTCAACAGACATGCTTAATGGTCT 246
QY 1180 GGAAGGAATATGACTGTGAACGGTGTTCAGATTGCCTTGTGCTTAAGTTAGGGCT 1239
DB 247 GGAAGGAATATGACTGTGAACGGTGTTCAGATTGCCTTGTGCTTAAGTTAGGGCT 306
QY 1240 GATGAAATGAAGTTATGACATGAGAAACTCTTGGCACTACAGATTCCTTAATG 1299

```

```

DB 307 GATGAAATGAAGTTATGACATGAGAACTCTTGGCACTACAGATTCCTTAATG 366
QY 1300 ACAGAGATTTCCTGTAATCTCAGGACATGACTAAATGGAAGTTGACTAAATGAG 1359
DB 367 ACAGAGATTTCCTGTAATCTCAGGACATGACTAAATGGAAGTTGACTAAATGAG 426
QY 1360 ATGAATCATAGCCAAAGTAAACGAGAAAGAAAGTTTCTACAGATTAGTCAGCCTGAG 1419
DB 427 ATGATGATAGCCAAAGTAAACGAGAAAGAAAGTTTCTACAGATTAGTCAGCCTGAG 486
QY 1420 GACACTAATGATGATGATGAGGACAGATGTTGTTGGATTGGCAGATCAGGCTTAGATT 1479
DB 487 GACACTAATGATGATGAGGACAGATGTTGTTGGATTGGCAGATCAGGCTTAGATT 546
QY 1480 AAAGGAACTTGCATTAGTGAAGTGAAGATGATGTTCTCCACTGTTATAGACACCA 1539
DB 547 AAAGGAACTTGCATTAGTGAAGTGAAGATGATGTTCTCCACTGTTATAGACACCA 606
QY 1540 GCAGCAAAATTCATCTAATGTTGTTGATTCCTATGG 1577
DB 607 GCAGCAAAATTCATCTAATGTTGTTGATTCCTATGG 644
RESULT 13
CB049545 654 bp mRNA linear EST 17-JAN-2003
LOCUS NISC_g111e08.y1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3271406
DEFINITION 5', mRNA sequence.
ACCESSION CB049545
VERSION CB049545.1 GI:277787832
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 654)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
cDNA Library Preparation:
cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: L1A8008 row: J column: 15
Seq primer: M13RPI reverse primer (ABI).
Location/Qualifiers
1..654
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3271406"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="NCI CGAP Pr28"
/note="Organ: prostate; Vector: pT7T3-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
ORIGIN

```

```

http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pyx-5.
Location/Qualifiers
1. 777
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6819584"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_EXO"
/note="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I;
Site_2: Not I; the library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:1791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGCAG. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."
```

3698	QY	GAUGGAGGAGAACUCCUAGGAGACACUCCACUATCCAGTGGGUGAUGATGACACAGAACGTT	3993
181	Db	GATGCGAGAGACCCCCAGGAAACAGATCCATCCAGTGGGTGGACGATGACAGACAGTT	240
3958	QY	AGCAAGGGTGTCTTAAGTCTCTATAGATGGGAAGTCCATGGAGACTATAAACAAATGTGAAG	4017
241	Db	AACAAAGGTGTGTGAGTCTCTATAGATGGGAAGTCTATGAATCTATAACAAAGCTGAAA	300
4018	QY	ATATTCATGGATCAGAAATATAAGCAAAATGGAAAAAGTAAATCAGATGGACAGAGTGTTT	4077
301	Db	ATATTCATGGGTCCAGATATAAAGCAAAACCGAAAAAGTCATCAGATGGACAGAGTGTTT	360
4078	QY	TTCTTAGAAAACCATGACACAGACAAATGCTCCTCAGATGATCCTCCAGATCACAGTAGATTG	4137
361	Db	TTCTTGAAAACCATGACCAACCAATACTGCTCGATGACCCCTGGAGATCACAGCAACTG	420
4138	QY	ACTGAGCATGTTGCCAAAGCTTTTTCGCTTGTCTCTGTGCTCCACCTGAAACTTCTGAAG	4197
421	Db	ACTGAGCATGTGCCAAGGCATTCTGCTCGCCCTCTGCCCCACCTGAAAGCTTTTAAAA	480
4198	QY	GAAGATGGAATGACAAACTGGGACTACGTGTGACACTTGCATCAGATCAGGTGGGTAT	4257
481	Db	GAAGATGGAATGACAAACTGGGACTCCGTGTGACCGCTTGACTCAGATCAGGTGGGTAT	540
4258	QY	CAAGCAGGAGCAATGSCAGGCCCTTCCCTCGCAGTACATGAATGATCTGGATAGCGCC	4317

Db 541 CAAGCAGGAGCAATGGCCAGCCCTTCCCTCGCAGTACATGACGATCTGCAGTGCC 600
 QY 4318 TTGGTGCCGGTGATCCATGAGGGGCGCTGCCAGCTTAGTGTAGGCGGCCGCTGTGTCATGAA 4377
 Db 601 TTGGTGCCGTGTCAATTCATGAGGAGCCCTGCCAGCTCAGTGAAGGCCCTGTGTCATGAA 660
 QY 4378 CTCACCTTTTATATCTCGAARAACATCGTATPAAACAGAGAAGACTTCATTTTCTGTT 4437
 Db 661 CTCATCTTTTATCTGGAARAACATCGCAT----AGATGAGGACTCCATGTTCTGTT 716
 QY 4438 CAGACTTGTGCAACAGCAGTGCATATCCCAAAATCAATTCACCTTTAAACTGGA 4490
 Db 717 NCAGACTGTTGAGCAGCAGTGCATATCCANATCAATTCACAATANACTGAA 769

RESULT 15
 BU745833
 LOCUS
 DEFINITION CH2#002 G03T7 Canine heart normalized cdna Library in pbluescript
 Canis familiaris cdna clone CH2#002_G03 5', mRNA sequence.
 ACCESSION BU745833
 VERSION BU745833.1 GI:23695416
 SOURCE EST..
 ORGANISM Canis familiaris (dog)
 Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 1 (bases 1 to 806)
 Yi, Y., Desai, R., Olarte, M., Henthorn, P. and George A.L.
 Expressed sequence tags from Canine heart
 Unpublished (2003)
 Other ESTs: CH2#002_G03T3
 Contact: George AL
 Division of Genetic Medicine
 Vanderbilt University
 529 Light Hall, 2215 Garland Avenue, Nashville, TN 37232-0275, USA
 Tel: 615 936 2660
 Fax: 615 936 2661
 Email: al.george@vanderbilt.edu
 Insert Length: 1688 Std Error: 0.00
 Seq primer T7: TAATACGACTCACTATAGG
 High quality sequence start: 51
 High quality sequence stop: 573.
 Location/Qualifiers
 1..806
 /organism="Canis familiaris"
 /mol_type="mRNA"
 /db_xref="taxon:9615"
 /clone="CH2#002_G03"
 /tissue_type="heart"
 /cell_type="heart"
 /dev_stage="mixed developmental stages (adult, 30 day - 40 day fetal)"
 /clone_lib="Canine heart normalized cdna Library in pbluescript"
 /note="Organ: heart; Vector: pbluescript; Site_1: 5' of vector NotI; Site_2: 3' of vector EcoRI; Tissue source: dog heart (adult, 30 day - 40 day fetal), right and left atria and ventricle. Dog breed - mixed (beagle, German shepherd, pointer, Irish setter). Library construction: oligo-dT primed"

FEATURES
source

Query Match 12.6%; Score 609.4; DB 13; Length 806;
 Best Local Similarity 88.8%; Pred. No. 5.1e-110;
 Matches 682; Conservative 0; Mismatches 82; Indels 4; Gaps 2;
 QY 1859 TTGGTGCGTGCAGACCCAGCAACCTTCTAACTTTAAACTTCAAATTCAAAGGCCATTAT 1918
 Db 34 TTGNGGTCTCGAACCAGCAACCTTCTAACTTTAAACTTCAAATTCATAGCCATTGT 93
 QY 1919 CAGACCAATTTCAAAATGACTTCTTCGCAACAGTGGAAATTAATAAATAAATG 1978

ORIGIN

Query Match 12.6%; Score 609.4; DB 13; Length 806;
 Best Local Similarity 88.8%; Pred. No. 5.1e-110;
 Matches 682; Conservative 0; Mismatches 82; Indels 4; Gaps 2;
 QY 1859 TTGGTGCGTGCAGACCCAGCAACCTTCTAACTTTAAACTTCAAATTCAAAGGCCATTAT 1918
 Db 34 TTGNGGTCTCGAACCAGCAACCTTCTAACTTTAAACTTCAAATTCATAGCCATTGT 93
 QY 1919 CAGACCAATTTCAAAATGACTTCTTCGCAACAGTGGAAATTAATAAATAAATG 1978

Search completed: May 18, 2004, 07:03:20
 Job time : 7424.14 secs

Db 94 CTGACCATTTGCAAAATGACCTTCTTACAAATCGTGGAAATTAATATAAATAAATAATG 153
 QY 1979 ATATTTCTTGGGAAGCAAAATAGGGGAAACTCAGCACCAATATGATCAGTCCATCTT 2038
 Db 154 ATGTTCTTGGGAAGCAAAATAGGGGAAACTCAGCACCAATATGATCAGTCCATCTT 213
 QY 2039 TGGGAACATCTCTAATGTCGATACAAATGGGAACATTTAGAAAGTTATAGAGGTGAGA 2098
 Db 214 TGGGAACATGTCGGTTGCTGATACAAATGTTTAAACATTTAGATACTACGAGTCTGGGA 273
 QY 2099 TCTCCACTAGACCATGCTTGCATAGCTCCAGATAGCCCAATATGATCTCAGAGCTG 2158
 Db 274 TATCCAGTAGACCGTGCCTTGCATAGCTCCAGAAAGCCCAATATGATCTCAGAGCTG 333
 QY 2159 GTCAGTTTGAATTTCTGCCAGAAAGCCCATTCACACGCTGGGTGAGGTGGCTCCAGTAT 2218
 Db 334 GTCATTTGGAATTTCTGCCAGAAAGCCCATTTACTACTCTGGGTGAGGTGGCTCCAGTAT 393
 QY 2219 GGATACCGGATTTCTAGGCTCCAAATTTGCATGAATGTGAAGCCAGGTTTACATTCACCA 2278
 Db 394 GGGTACCAGATTTCTAGGCTCCAAATTTGCATGAATGTGAAGCCAGGTTTACATTCACCA 453
 QY 2279 AAGGAGGATCACTGCAGAGCATGTGGGAAGGTTTTCTGCTTCTCTGCTGTAGCCTGA 2338
 Db 454 AAGGAGGATCACTGCAGAGCATGTGGGAAGGTTTTCTGCTTCTCTGCTGTAGCCTGA 513
 QY 2339 AATGTAACCTGTTTATACATGGACAGAAAGAAAGCTAGAGTGTGTAATCTGCCATTCAG 2398
 Db 514 AATGTAACCTGTTTATACATGGATAGAAAGAGGCTAGAGTGTGTAATCTGCCATTCAG 573
 QY 2399 TGCTAATGATGCTCAAGCTTGGGAGAAATGATGATGCTCAAGCCAGAGCCCTTAACC 2458
 Db 574 TGCTAATGATGCTCAAGCTTGGGAGAAATGATGATGCTCAAGCCAGAGCCCTTAACC 633
 QY 2459 CTAAACATCTGCTGAATACTGTTTCTACTATCCCTTTCAGCAAGCTCAGGCTTCAG 2518
 Db 634 CTAAACATCTGCTGAATACTGTTTCTACTATCCCTTTCAGCAAGCTCAGGCTTAAG 693
 QY 2519 GAGCTCTGAGCTCTCCACCTCCCACTGTGATGCTGAGGAGTTTAAAGCACCCTG 2578
 Db 694 GAACCTTGAGCTCTCCACCTCCCACTGTGATGCTGAGGAGTTTAAAGCACCCTG 753
 QY 2579 GAG---CAGAGTGGCTCAGCCC-AGAGAGCAGAGCGAGTTTGGTTT 2622
 Db 754 TGGAGCAAGATGTGGCTCAACCCAAAGACAGAGCGGATTTTGGTT 801